S

SAR INTERFEROMETRY

Masato Furuya
Department of Natural History Sciences, Hokkaido University, Sapporo, Japan

Synonyms

Differential InSAR (abbreviated as D-InSAR); Interferometric SAR (abbreviated as InSAR); Radar interferometry; SAR interferometry

Definition

Radar. Acronym standing for Radio Detection and Ranging. A technique to detect any targets and measure the distance to them, based on the round-trip time of microwave (radio wave) pulses between the antenna and the targets.
SAR. Acronym standing for Synthetic Aperture Radar. A technique to image any ground surfaces, using airborne or spaceborne radar sensor. Its high spatial resolution is achieved by collecting numerous return pulses from each target in sight and by effectively synthesizing large antenna size.
InSAR. Acronym standing for Interferometric SAR. A technique to image surface topography and ground displacements, using phase values of two or more SAR images.

Introduction

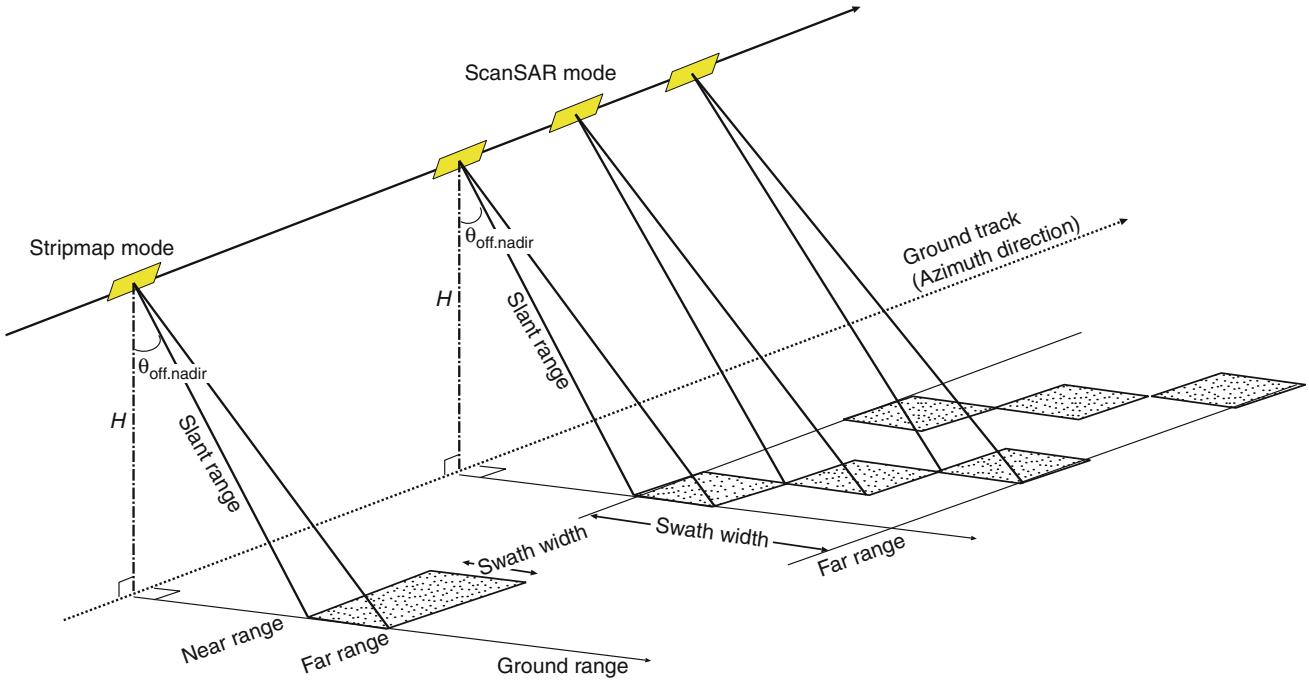
Crustal deformation data have been traditionally acquired by ground-based geodetic techniques such as leveling, triangulation, and electro-optic distance measurement. More recently, global positioning system (GPS) has become a standard tool for high-precision crustal deformation measurement, and provided us with a wealth of data to study plate tectonics, earthquakes, volcanic activities, and atmospheric and hydrological loading deformation. All these techniques, however, require in situ benchmarks, and thus prevent us from observing inaccessible areas. Interferometric SAR (InSAR) was, therefore, regarded as a surprising and revolutionary technique when Massonnet et al. (1993) first showed an image of the co-seismic deformation associated with the 1992 M7.3 Landers earthquake, because the raw data was completely acquired on a spaceborne sensor. Another big surprise for the community was its incredibly high spatial resolution, which no other geodetic techniques were possible to achieve in practice.

Nowadays, InSAR users have proliferated in a worldwide community and applied to a variety of geophysical problems. A number of excellent and extensive reviews for advanced students and researchers are already published (Bamler and Hartl, 1998; Massonnet and Feigl, 1998; Burgmann et al., 2000; Hanssen, 2001; Pritchard, 2006; Simons and Rosen, 2007; Zhou et al., 2009). I therefore tried to make this article much shorter and more introductory, but it still includes necessary and useful concepts, ranging from the fundamentals of SAR/InSAR imagery to more up-to-date topics.

Fundamentals of SAR imaging and SAR data

SAR satellite flies over at an altitude of hundreds of km, repeating transmission and reception of microwave pulses. The along-track and across-track axes are almost identical to the *azimuth* and *range* axis in the acquired radar image. The area illuminated on the ground is called *swath*, whose width spans roughly 50–100 km in the standard *stripmap* (or *strip*) mode with an incidence angle of 20–50° (Figure 1). While previous SAR applications are mostly derived from the stripmap mode, another imaging mode, *ScanSAR*, is also promising because it covers much wider swath width, 300–500 km, by illuminating multiple swaths at the expense of reducing the resolution. ScanSAR is useful for imaging long-wavelength signals

Harsh K. Gupta (ed.), *Encyclopedia of Solid Earth Geophysics*, DOI 10.1007/978-90-481-8702-7,



SAR Interferometry, Figure 1 Geometry of SAR imaging. SAR sensor transmits microwave pulses in slant range direction, and receives their reflected pulses. While stripmap mode achieves high spatial resolution with a fixed off-nadir angle, ScanSAR mode achieves wider imaged area (swath) with multiple off-nadir angles at the expense of the resolution.

associated with, for instance, a magnitude-8-class earthquake (Motagh et al., 2008).

Although it was not strictly necessary, satellite-based SAR system has been often placed on a sun-synchronous and near-polar orbit with an inclination angle of slightly greater than 90°. When the satellite moves to the north (south), we call it is in *ascending* (*descending*) orbit.

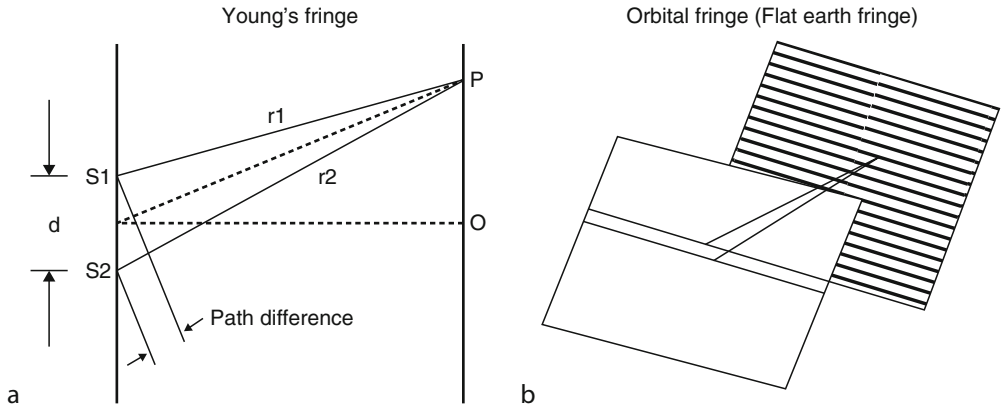
The raw data acquired on SAR sensor are impossible to visually interpret, and require a bit involved processing algorithms; those algorithms are detailed in a couple of text books (e.g., Curlander and McDonough, 1991; Cumming and Wong, 2005). The first interpretable SAR image is a single-look-complex (SLC) image, whose important difference from other optical images is that each pixel consists of a complex (real and imaginary) value, i.e., amplitude and phase. This is because the waveform of each repeated pulse is precisely controlled to be identical, and hence the received pulse provides us with not only a scattering (reflection) *intensity* but also a *phase*. The phase data do contain the geometric information from the antenna to the ground targets, and are fully exploited in generating InSAR image as discussed later. However, the phase image itself is usually not as useful as the intensity image because it is again impossible to visually interpret the physical meaning. Meanwhile, the intensity image is often useful and derived from a square-root magnitude of SLC data with spatial averaging called *multi-looking*. By single-look, it means the finest spatial resolution for both range and azimuth axis. In the standard stripmap mode, the range and azimuth resolutions are derived as,

$$\Delta r = \frac{c}{2B}, \quad and \quad \Delta a = \frac{L}{2}, \tag{1}$$

respectively; the c, B, and L are the speed of light, the frequency bandwidth of the microwave pulse, and the antenna length along azimuth axis, respectively (Curlander and McDonough, 1991; Cumming and Wong, 2005). The waveform of each microwave pulse is called *chirp signal*, whose instantaneous frequency linearly changes by as much as the frequency bandwidth B over the duration of each pulse. It should be noted that the spatial resolution depends neither on the sensor altitude nor the carrier frequency of microwave. Intensity images are often shown in gray scale images, in which strongly (weakly) reflected objects/areas are usually colored as bright (dark). Although they simply look like black-and-white photographs, we should keep in mind that they could be acquired regardless of weather and time because SAR is actively transmitting and receiving microwaves. Also, intensity images are indispensable for high-precision image matching prior to a generation of InSAR image.

Fundamental principles of InSAR

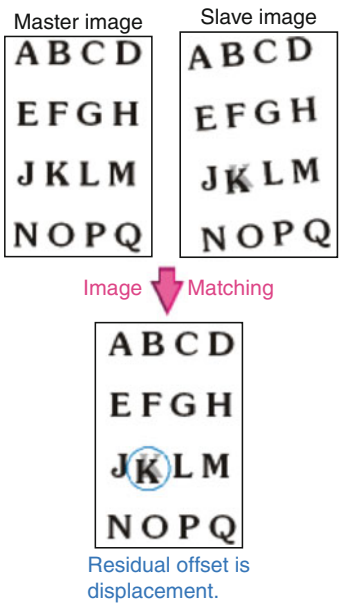
Interferometric SAR (InSAR) is a technique to generate a digital elevation model (DEM) or a ground displacement



SAR Interferometry, Figure 2 (**a**) Geometry of the Young's experiment. Depending on the path difference, the two coherent waves from the slit, S1 and S2, are in-phase or out-of-phase on the screen, and interference fringes are observed on the right screen. (**b**) Orbital fringe (flat earth fringe) can be regarded as a 3-D analogue of the Young's experiment.

image from a *pair* of SLC images. The term *interferogram* is often used to represent InSAR image. We can understand the principle of InSAR, recalling the classical Young's experiment that is known to be a proof of the wave characteristics of the light (Ghilia and Pritt, 1998). Two coherent waves out of the slits will generate "stripes" on the wall, called interference *fringe* (Figure 2a). We can simulate the fringe if we know the separation of the slits, the distance from each slit to the wall, and the wavelength of the coherent wave. Depending on the path difference, the two coherent waves are in-phase or out-of-phase when they reach the screen. Namely, the difference of the phases generates the interference fringe. We may regard the imaging geometry of InSAR as the 3-D Young's experiment (Figure 2b). The repeat orbit tracks, the ground surface, and the microwave correspond to the double slits, the screen, and the coherent wave, respectively. Once we get two SLC images, we can generate an initial interferogram, multiplying one SLC image with the complex conjugate of the other SLC image. We then observe similar fringes in the initial interferogram as illustrated in Figure 2b, which is literally a map of the difference of two SLC phases. For descriptive purposes, the former SLC image is often denoted as *master*, and the latter SLC image is called *slave*. At this moment, the slave image must be precisely co-registered (or matched) to the master image (Figure 3); we will come back to this *image co-registration* (or *image matching*) procedure later on.

While Figure 2b shows an initial interferogram over flat areas with parallel orbits, the fringe will appear undulated if the areas are not flat. The fringe over flat areas is called *flat Earth fringe* (or, *orbital fringe*), and can be precisely simulated from the pair of orbit data. If we subtract the flat Earth fringes from the initial interferogram, we can extract *topographic fringe* that can be used to generate DEM. The Shuttle Radar Topography Mission (SRTM) was carried out along this idea in 2001, and generated 3-s resolution DEM over +/− 60° latitudes (Farr et al., 2007). In the case



SAR Interferometry, Figure 3 Image registration (matching) of the master and slave images prior to interferogram generation, and the principle of pixel-offset technique to derive large displacements. While long-wavelength distortion can be corrected, localized huge displacement remains as residual offset. Courtesy of Tobita et al. (2001a).

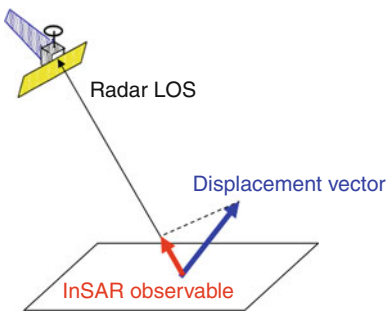
of SRTM, they carried two SAR antennas on the same platform, and thus were able to generate DEM without repeating the previous orbit track. In contrast, all the present SAR satellite systems carry only one antenna with a repeat-pass period of several weeks, which are 11 days for TerraSAR-X, 16 days for COSMO-SkyMed, 24 days for Radarsat-1/2, 35 days for Envisat, and 46 days for ALOS. Therefore, if ground surface undergoes significant deformation during the repeat orbit cycles due, for

instance, to earthquake and volcanic eruption, the interferogram will include *deformation fringe* as well. To extract deformation fringe, we must take out both orbital fringe and topographic fringe, which can be simulated from satellite orbit data and DEM. The deformation fringes represent slant range changes along the radar line-of-sight (LOS), and thus projections of the 3-D displacement vector on the ground along the unitary vectors toward the radar LOS (Figure 4). The range changes should be interpreted as relative displacements to the reference point(s) inside each interferogram. Depending on literatures, they denote differential interferometric SAR (D-InSAR) when the technique is used to detect deformation signals. Recently, however, the term InSAR is often and simply used to represent D-InSAR.

Even if no significant ground displacements take place during the repeat-pass period, however, we usually encounter other non-negligible fringes due to the spatial heterogeneities in the propagation delay of microwaves through the atmosphere, the errors in satellite orbit data, and those in DEM. Because these fringes limit the precision and accuracy of SAR-based crustal deformation measurement, a couple of correction approaches have been proposed. More advanced time-series analysis techniques have also been developed to overcome the issues, which will be introduced in the last section.

InSAR processing

Image registration (Matching): Before we get an initial interferogram, we must register (or, match) each imaged target in one SLC image to the same target in the other SLC image with a sub-pixel level accuracy, because any ground objects do not usually locate at the same pixel coordinates in each SLC image. This pre-processing is called *image registration* (or *image matching*) and prerequisite to be performed prior to generating an initial interferogram. Although a simple polynomial transformation between the range and azimuth coordinates of two SLC images is sufficient in most cases, we need to take into account the effects of 3-D topography when the terrain surface is rugged to eliminate a stereoscopic effect (Michel et al., 1999).



SAR Interferometry, Figure 4 InSAR observable is a projection of the displacement vector along the radar line-of-sight (LOS) direction.

When large ground displacements on the order of meters or more take place locally, and if we correct for the long-wavelength image distortion using the polynomial transformation, we can detect and quantify those localized displacements as a by-product of image registration without viewing InSAR image (Figure 3; Tobita et al., 2001a). This approach to detect large displacements is called *pixel offset* or *feature tracking* technique, and has been applied to earthquakes, volcanic eruptions, and glacier movements. The advantages of pixel-offset data are twofolds. First, pixel-offset data can quantify large displacements even in such areas that completely loses interferometric coherence, where InSAR data cannot be unwrapped; we describe *coherence* and *unwrapping* later below. Secondly, in contrast to InSAR data, pixel-offset data provide us with not only range offset but also azimuth offset component. While the range offset has the same sensitivity to the 3-D displacement vector as InSAR data (Figure 4), the azimuth offset is a projection of the displacement vector onto the unitary vector perpendicular to the LOS. Hence, the azimuth offset data are complementary to the range offset or InSAR data. Taking advantage of this property, Fialko et al. (2001) derived a full 3-D displacement map for the 1999 M7.1 Hector Mine earthquake, combining the InSAR data from both ascending and descending track with the azimuth offset data. Using pixel-offset data from both descending and ascending track, Tobita et al. (2001a, b) inferred a 3-D displacement map associated with the 2,000 eruption episode at Usu volcano.

Interferometric phase and its relation to geometry: Suppose we have two co-registered SLC images, E_1 and E_2, acquired from different ranges r_1 and r_2:

$$E_1 = e^{j\phi_{Scatter}} e^{-\frac{4\pi r_1}{\lambda}} \tag{2a}$$

$$E_2 = e^{j\phi_{Scatter}} e^{-\frac{4\pi r_2}{\lambda}} \tag{2b}$$

Here we assume that the reflection magnitude and scattering phase are constant during the data acquisition time. Then, the interferometric phase ϕ is derived as

$$E_1 E_2^* = e^{-\frac{4\pi(r_1 - r_2)}{\lambda}} \tag{3}$$

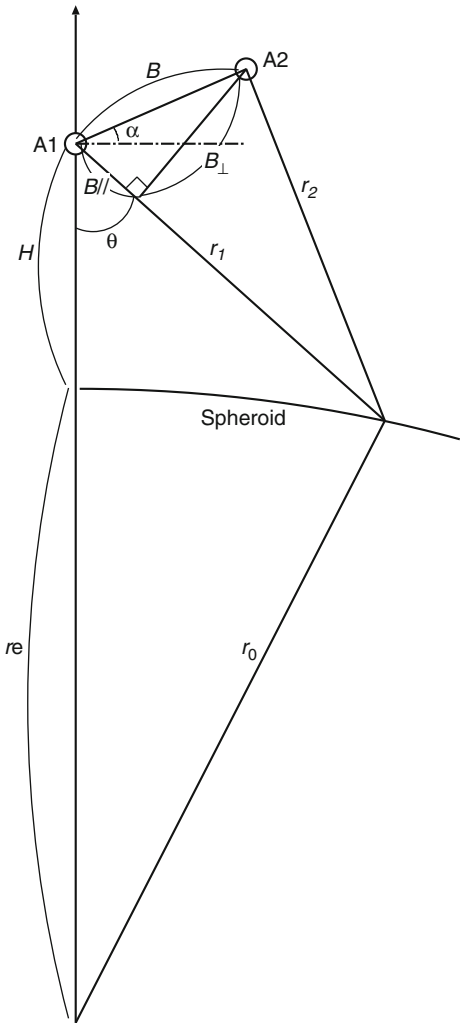
or

$$\phi = \frac{4\pi}{\lambda}(r_1 - r_2) \tag{4}$$

The last one is the fundamental equation for InSAR, which describes "unwrapped" phase in the initial interferogram. The actual phase in the initial interferogram is "wrapped" into an interval $[-\pi, \pi]$, and thus has ambiguities of $2\pi N$; N is integer. In order to quantify the ground displacement along radar LOS, we have to perform 2-D *phase unwrapping* on the interferogram, which is not necessarily straightforward (Bamler and Hartl, 1998; Ghilia and Pritt, 1998). While the interferometric phase is strictly a phase "difference" of two SLC phases, it is conventional

to simply call phase. The factor 4 is to take into account the round-trip distances.

Figure 5 is a cross section that is perpendicular to the satellite repeat tracks and passes through the Earth's center, and shows a geometry of InSAR data acquisition. The spatial separation of the repeating orbits is termed *baseline* (or *spatial baseline*), B; the temporal separation of data acquisition is sometimes called *temporal baseline*. Because the baseline B is usually much shorter than the ground range distance R, a parallel ray approximation holds (Zebker et al., 1994) and the fundamental Equation 4 can be approximated as follows:



SAR Interferometry, Figure 5 Geometry of InSAR data acquisition and its relation to the baseline. The A1 and A2 are the satellite's repeat orbits, and the spatial distance between the A1 and A2 is the baseline B. The initial InSAR phase is proportional to the difference between the ranges, r_1 and r_2, and hence the B_{para} (eq. 5). The fringe rate (density) along the range axis is proportional to the B_{perp} (eq. 6).

$$\phi = \frac{4\pi}{\lambda}(r_1 - r_2) \approx -\frac{4\pi}{\lambda}B_{//} = B\sin(\theta - \alpha), \qquad (5)$$

where θ and α are defined in Figure 5, and $B_{//}$ (or B_{para}) is a baseline component parallel to the radar LOS. The angle θ is called *off-nadir angle*, and is identical to incidence angle if the Earth's curvature is negligible. The other baseline component $B_\perp$ (or B_{perp}) is perpendicular to radar LOS and gives us an important criterion for successful InSAR processing as we discuss below.

Decorrelation (Coherence): In the actual InSAR data processing, we do not necessarily get clear fringes over the entire area. Depending on the data pairs and places, it is not uncommon that no fringes are observed. To detect clear fringes, the reflected waves received at master and slave acquisitions must be more or less correlated to each other. The degree of correlation is quantified as *coherence*, and there are two independent decorrelation sources.

The first source of decorrelation originates in the imaging geometry. As Figure 6 indicates, we observed higher (fewer) fringe density as becomes longer (shorter); imagine the case of zero baseline length. The fringe density can be derived from the gradient of phase (Equation 5) along the range axis:

$$\frac{\partial\phi}{\partial R} = -\frac{4\pi B_\perp}{\lambda R\tan\theta} + \frac{4\pi B_\perp}{\lambda(r_e + H)\sin\theta} \approx -\frac{4\pi B_\perp}{\lambda R\tan\theta}. \qquad (6)$$

Namely, the fringe density is proportional to the perpendicular baseline $B_\perp$, and inversely proportional to the wavelength λ; see Simons and Rosen (2007) for the case with topography. If the fringe density becomes too high to be counted within a range resolution of SAR image, we will not be able to identify any orbital fringes. This type of decorrelation is termed *baseline decorrelation* (or *spatial decorrelation*). The *critical baseline* is given as such a perpendicular baseline that gives a phase gradient 2π over the range resolution Δr;

$$B_\perp^c = \frac{\lambda R\tan\theta}{2\Delta r}.$$

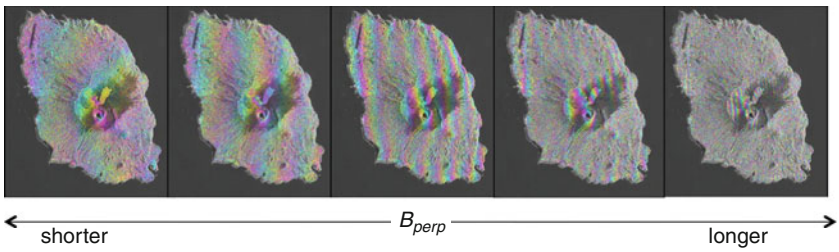
For a typical value of ALOS/PALSAR with $\lambda = 23(cm)$, $R = 870(km)$, $\theta = 34°$, $\Delta r = 5(m)$, the critical baseline becomes $B_\perp^c = 135,000(m)$, which gives an upper limit of $B_\perp$. However, we practically prefer much shorter $B_\perp$, generally less than ~2,000 m for ALOS/PALSAR, because in more realistic situations the effect of topography also comes in. The longer the $B_\perp$, the more sensitive to rugged terrain as Figure 6 indicates. To eliminate topographic fringes, we need more accurate and higher resolution DEM if the $B_\perp$ becomes longer. Massonnet et al. (1996) proposed an alternative approach that could effectively reduce the $B_\perp$ by a combination of integer multiplied (wrapped) interferograms. For instance, if one interferogram with perpendicular baseline of 300 m is combined with the other interferogram with perpendicular baseline of 290 m with factors 1 and −1, the effective perpendicular baseline becomes 10 m.

The scaling operation, however, also scales the amount of noise, and thus the approach is limited to small integer numbers.

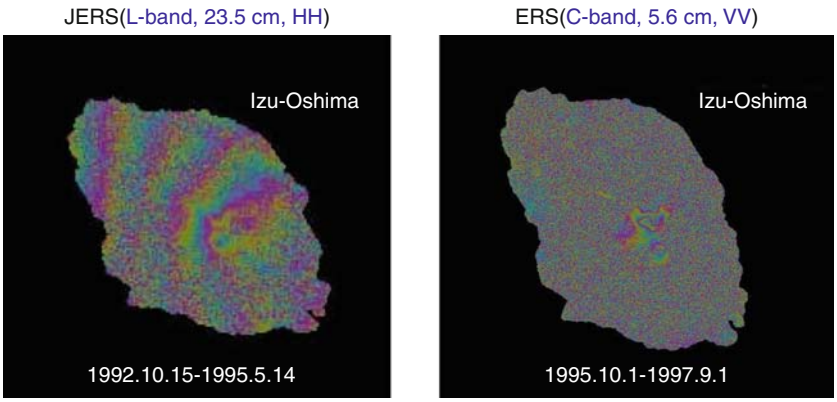
The second type of decorrelation is termed *temporal decorrelation*, which is related to the scattering phase in the Equation 2a, and originates in how the microwave pulses interact with the physical objects near the ground. We often encounter the temporal decorrelation problem over vegetated areas with C-band (shorter-wavelength) SAR data and/or snow-covered areas; see Figure 7. It should be recalled that each pixel value in SLC image is a superposition of all the reflected microwaves from all scatterers inside each resolution cell ($\sim 5 \times \sim 10$ m). Short-wavelength microwave pulses tend to be reflected on the vegetation canopies before reaching the ground surface, and their random motion will result in different scattering phases at different acquisition time, causing temporal decorrelation. On the contrary, long-wavelength microwave pulses can more easily reach the ground, which does not move as rapidly as vegetations, and thus the resulting scattering phases will be also stable over time. Besides the selection of wavelength, the polarization of microwave is also essential for better coherence over time. While, most presently, operated satellite-SAR sensors are capable of multi-polarization modes, it was shown that HH-polarization gives better coherence than VV-polarization (Cloude and Papathanassiou, 1998). This is because the HH-polarized pulses can more easily penetrate through vegetations.

Outlook for InSAR geodesy

Limitations of present InSAR: Although it has a potential to detect tens of km-scale or even larger-scale secular deformation signals on the order of mm/year, InSAR technique has been most successfully applied to detection of spatially localized signals on the order of centimeters or more, such as those associated with earthquakes, volcanic eruptions, and ground subsidence. This is because the artifacts due to inaccurate satellite orbit data and/or microwave propagation delays (advances) in the troposphere (ionosphere) can mask small-amplitude, long-wavelength deformation signals that are similar in both their amplitude and the spatial scale.



SAR Interferometry, Figure 6 The fringe rate (density) depends on the B_{perp}; see eq (6). The shorter the B_{perp}, the fewer the observed fringes, and thus better to detect deformation signals. In order words, there is a limit in the B_{perp} over which we cannot count the number of fringes. The InSAR image is based on JERS data over Izu-Oshima volcano island, Japan. Original SAR data is copyrighted by JAXA and MITI, Japan.



SAR Interferometry, Figure 7 Comparison of two interferograms at Izu-Osima volcano, derived from (left) L-band HH JERS data and (right) C-band VV ERS data. While clear fringes are observed to the left even with 2.5 years temporal baseline, we can recognize the fringes only around the caldera that are covered with few vegetations.

Although high-precision orbit data are indispensable to correct for the orbital fringes in the initial interferograms, their errors even on the order of 10 cm or less will generate non-negligible long-wavelength artifacts, which usually look like curved surfaces in the entire interferogram (e.g., Hanssen, 2001). Conventionally, they are fitted with low-order polynomials and simply taken out unless any sort of stacking or time-series analysis discussed below is applied. While this procedure works to eliminate the artifacts due to orbit errors, it will also take out any long-wavelength geophysically interesting signals such as the inter-seismic, post-seismic, ocean tidal loading, solid-Earth tide, and post-glacial rebound signals. Alternatively, if the ground control points (GCP) are available, where the precision ground deformation data are available, we can reestimate the baseline, based on those GCP data (e.g., Rosen et al., 1996), but such data are often unavailable in remote areas.

One approach to correct for the tropospheric delay signals is to employ the other independent estimates derived from either the GPS-based tropospheric delay estimates (e.g., Onn and Zebker, 2006) or the output results from high-resolution numerical weather model (e.g., Foster et al., 2006). These so-called calibration approaches are, however, not easily applicable. The dense ground-based GPS network is limited to a few areas in the world. Also, high-resolution numerical weather model still needs significant computational resources.

Besides the tropospheric delay problem, the effects of ionosphere on both interferograms and pixel-offset images were clearly recognized in the results of the 2008 Wenchuan earthquake (Mw7.9), China, derived from ALOS/PALSAR (Kobayashi et al., 2009; Raucoules and de Michele, 2010), although they were pointed out in polar region many years ago (e.g., Matter and Gray, 2002). It is well known that the lower the carrier frequency is, the more significant the ionospheric dispersion impacts on the propagation delay. Thus, in many of the previous applications of C-band SAR data, the effects of ionosphere could have been neglected. While GPS also employs L-band, the high-precision GPS geodetic survey corrects for the ionospheric effect with the use of dual frequency, L1 and L2, observation data. In contrast, PALSAR is a single frequency SAR sensor and incapable of the standard ionosphere-correction approach. Empirically, however, we will encounter the ionospheric signals more frequently in the ascending data acquired in the local nighttime than in the descending data acquired in the local daytime. We also recall that the JERS, the other L-band SAR operated during 1992–1998, did not reveal any significant ionospheric signals at least in mid-latitude regions, and that most of the JERS data were acquired in the descending track. Besides the latitude, the effects of ionosphere on SAR image might, therefore, significantly depend on the data acquisition time. Like the tropospheric effects, detailed studies of ionospheric impacts on the SAR data are also currently underway.

A simple approach to eliminate those noises is *stacking*, in which several interferograms are stacked to isolate small-amplitude signals, because those noises can be regarded as temporally random, whereas the deformation signals are spatially persistent. Two important prerequisites for successful stacking are: (1) the data acquisition dates of those interferograms should not be overlapped, in order not to enhance the noises of any particular acquisition date(s), and (2) each temporal baseline should be as long as possible so that each interferogram can include as much deformation signals as possible. In reality, it is not easy to gather many independent interferograms that have desirably long temporal baselines because the available data often encounter the spatial and temporal decorrelation. Also, the simple stacking approach inherently assumes temporally linear evolution in the ground deformation, preventing us from deriving time-series data.

Time-Series Analysis: Ferretti et al. (2000, 2001) proposed a new analysis technique called *Permanent Scatterer InSAR* (*PS-InSAR*), in which they take advantage of even such data pairs whose spatial baselines are longer than the critical values. Thereby, they could expand the temporal coverage, and thus could estimate the long-term deformation signals on the order of mm/year. Key idea of PS-InSAR is to pick up only such pixels that will exhibit long-term coherence due to the existence of corner-reflector-like targets, which Ferretti et al. called "permanent scatterers." Based on those pixels alone, they generate a stack of differential interferograms, using available DEM and orbit data. The phase values include not only deformation signals, but also such topographic signals that were not initially taken into account, because the longer spatial baseline pairs are so sensitive to the topography that the available DEM could not account for. In PS-InSAR and its variants (Werner et al., 2003; Hooper et al., 2004), they fit the differential interferogram stack to a phase model that describes not only temporal evolution of deformation but also corrections to the available DEM. Deviations from the phase model can be filtered into either non-linear deformation or atmospheric signals because the former signals are correlated and thus low-pass filtered along temporal axis, while the latter signals are temporally random; the orbit data must be assumed to be correct. A known limitation of PS-InSAR is its rather lower sampling density over non-urban areas. However, despite a lack of man-made objects, Furuya et al. (2007) succeeded in detecting active salt tectonic motion, applying a similar technique to Canyonlands National Park, Utah, presumably because the area was non-vegetated and the exposed surface rocks behaved like corner-reflector-like targets.

Another time-series analysis approach was devised and known as small baseline subset (*SBAS*) algorithm (Berardino et al., 2002). Key idea of the SBAS algorithm is *least-squares inversion* of unknown deformation at each SAR data acquisition epoch, based on the available unwrapped differential interferograms (e.g., Lundgren

et al., 2001; Schmidt and Burgmann, 2003). Using small baseline interferometric pairs, the SBAS approach is free from spatial decorrelation and allows us to take advantage of the fine spatial resolution of InSAR data. If the number of interferograms is greater than or equal to the number of SAR acquisitions, the inversion problem becomes an over-determined or well-determined problem, and can be easily solved in a least-squares approach. It is uncommon, however, that all the available interferometric pairs have short baselines, and accordingly the temporal sampling rate will decrease. Berardino et al. (2002) proposed to employ several groups of "small baseline subset" to overcome the lower temporal resolution issue, and solved the rank-deficient problem with the use of singular value decomposition (SVD) technique. The SVD gives the minimum-norm least-squares solution, which is equivalent to minimizing the estimated velocities at any time intervals.

Time-series analysis of SAR data is a promising technique, but almost all previous analyses are based on the C-band ERS1/2 and Envisat data, because not only more-than-decade-long data but also high-precision, well-controlled satellite orbits are available for these satellites. As noted before, not all geophysically interesting phenomena could be detected by C-band and shorter-wavelength SAR data. If the L-band ALOS/PALSAR data are archived for a much longer time, and if the follow-on ALOS-2 and the DESDynI are launched as scheduled, the time-series analysis of SAR data will become feasible even in areas that have never been monitored before. The time-series analysis with ScanSAR data should also be possible. Long-term continuous monitoring with L-band SAR will provide us with more opportunities for new discoveries.

Bibliography

Bamler, R., and Hartl, P., 1998. Synthetic aperture radar interferometry. *Inverse Problems*, **14**, R1.

Berardino, P., Fornaro, G., Lanari, R., and Sansosti, E., 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience Remote Sensing*, **40**, 2375.

Burgmann, R., Rosen, P. A., and Fielding, E. J., 2000. Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Annual Review of Earth and Planetary Sciences*, **28**, 169.

Cloude, S. R., and Papathanassiou, K. P., 1998. Polarimetric SAR Interferometry. *IEEE Transactions on Geoscience Remote Sensing*, **36**, 1551.

Cumming, I. G., and Wong, F. H., 2005. *Digital Processing of Synthetic Aperture Radar Data: Algorithm and Implementation*. Boston: Artech House.

Curlander, J. C., and McDonough, R. N., 1991. *Synthetic Aperture Radar: Systems and Signal Processing*. New York: Wiley interscience.

Farr, T. G., et al., 2007. The shuttle radar topography mission. *Reviews of Geophysics*, **45**, RG2004.

Ferretti, A., Prati, C., and Rocca, F., 2000. Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 2202.

Ferretti, A., Prati, C., and Rocca, F., 2001. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 8.

Fialko, Y., Simons, M., and Agnew, D., 2001. The complete (3-D) surface displacement field in the epicentral area of the 1999 Mw7.1 hector mine earthquake, California, from space geodetic observations. *Geophysical Research Letters*, **28**, 3063.

Foster, J., Brooks, B., Cherubini, T., Shacat, C., Businger, S., and Werner, C. L., 2006. Mitigating atmospheric noise for InSAR using a high resolution weather model. *Geophysical Research Letters*, **33**, L16304.

Furuya, M., Mueller, K., and Wahr, J., 2007. Active salt tectonics in the Needles District, Canyonlands (Utah) as detected by interferometric synthetic aperture radar and point target analysis: 1992–2002. *Journal of Geophysical Research*, **112**, B06418.

Ghilia, D. C., and Pritt, M. D., 1998. *Two Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. New York: Wiley.

Hanssen, R. F., 2001. *Radar Interferometry: Data Interpretation and Error Analysis*. Dordrecht: Kluwer.

Hooper, A., Zebker, H., Segall, P., and Kempes, B., 2004. A new method for measuring deformation on volcanos and other natural terrains using InSAR persistent scatterers. *Geophysical Research Letters*, **31**, L23611.

Kobayashi, T., Takada, Y., Furuya, M., and Murakami, M., 2009. Location and types of ruptures involved in the 2008 Sichuan Earthquake inferred from SAR image matching. *Geophysical Research Letters*, **36**, L07302.

Lundgren, P., Usai, S., Sansosti, E., Lanari, R., Tesauro, M., Fornaro, G., and Berardino, P., 2001. Modeling surface deformation observed with synthetic aperture radar interferometry at Campi Flegrei caldera. *Journal of Geophysical Research*, **106** (B9), 19355.

Massonnet, D., and Feigl, K. L., 1998. Radar interferometry and its application to changes in the earth's surface. *Reviews of Geophysics*, **36**, 331.

Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Raboute, T., 1993. The displacement field of the Landers earthquake mapped by radar interferometry. *Nature*, **364**, 138.

Massonnet, D., Vadon, H., and Rossi, M., 1996. Reduction of the need for phase unwrapping in radar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 489.

Matter, K. E., and Gray, A. L., 2002. Reducing ionospheric electron density errors in satellite radar interferometry applications. *Canadian Journal of Remote Sensing*, **28**, 583.

Michel, R., Avouac, J.-P., and Taboury, J., 1999. Measuring ground displacements from SAR amplitude images: application to the landers earthquake. *Geophysical Research Letters*, **26**, 875.

Motagh, M., Wang, R., Walter, T. R., Bürgmann, R., Fielding, E., Anderssohn, J., and Zschau, J., 2008. Coseismic slip model of the 2007 august pisco earthquake (Peru) as constrained by wide swath radar observations. *Geophysical Journal International*, **174**, 842.

Onn, F., and Zebker, H. A., 2006. Correction for interferometric synthetic aperture radar atmospheric phase artifacts using time series of zenith wet delay observations from a GPS network. *Journal of Geophysical Research*, **111**, B09102.

Pritchard, M. E., 2006. InSAR, a tool for measuring Earth's surface deformation. *Physics Today*, **59**(7), 68.

Raucoules, D., and de Michele, M., 2010. Assessing Ionospheric Influence on L-Band SAR Data: Implications on Coseismic Displacement Measurements of the 2008 Sichuan Earthquake. *IEEE Geoscience Remote Sensing Letters*, **7**, 286.

Rosen, P. A., Hensley, S., Zebker, H. A., Webb, F. H., and Fielding, E. J., 1996. Surface deformation and coherence measurements of

Kilauea volcano, Hawaii, from SIR-C radar interferometry. *Journal of Geophysical Research*, **101**(E10), 23109.
Schmidt, D. A., and Burgmann, R., 2003. Time-dependent land uplift and subsidence in the Santa Clara valley, California, from a large interferometric synthetic aperture radar data set. *Journal of Geophysical Research*, **108**(B9), 2416.
Simons, M., and Rosen, P. A., 2007. Interferometric synthetic aperture radar geodesy. In Herring, T. A. (ed.), *Treatise on Geophysics*. New York: Elsevier, Vol. 3, pp. 391–446.
Tobita, M., Murakami, M., Nakagawa, H., Yarai, H., and Fujiwara, S., 2001a. Two-dimensional field of three-dimensional components of deformations and velocities, and volume change around Usu Volcano associated with the 2000 eruption by matching of SAR images (in Japanese). *Journal of Geographical Survey Institute*, **95**, 37.
Tobita, M., Murakami, M., Nakagawa, H., Yarai, H., Fujiwara, S., and Rosen, P. A., 2001b. 3D surface deformation of the 2000 Usu Eruption measured by matching of SAR images. *Geophysical Research Letters*, **28**, 4291.
Werner, C. L., Wegmuller, U., Strozzi, T., and Wiesmann, A., 2003. Interferometric point target analysis for deformation mapping, paper presented at IGARSS'03, Toulouse: Geoscience Remote Sensing Society.
Zebker, H. A., Rosen, P. A., Goldstein, R. M., Gabriel, A., and Werner, C. L., 1994. On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. *Journal of Geophysical Research*, **99**(B10), 19617–19634.
Zhou, X. B., Chang, N. B., and Li, S. S., 2009. Applications of SAR Interferometry in Earth and Environmental Science Research. *Sensors*, **9**, 1876.

Cross-references

Earthquakes and Crustal Deformation
Earthquake Rupture: Inverse Problem
Earthquakes, Source Theory
Geodesy, Ground Positioning and Leveling
GPS, Tectonic Geodesy
Inverse Theory, Global Optimization
Inverse Theory, Linear
Inverse Theory, Monte Carlo Method
Inverse Theory, Singular Value Decomposition
Remote Sensing and GIS Techniques for Tectonic Studies
Remote Sensing, Applications to Geophysics
Slow Earthquake

SATELLITE LASER RANGING

David Coulot[1], Florent Deleflie[2], Pascal Bonnefond[3], Pierre Exertier[3], Olivier Laurain[3], Bertrand de Saint-Jean[3]
[1]Laboratoire de Recherche en Géodésie, Institut Géographique National, GRGS Ecole Nationale des Sciences Géographique, Marne la Vallée Cedex 2, France
[2]Institut de Mécanique Céleste et de Calcul des Ephémérides, GRGS, Paris, France
[3]Observatoire de la Côte d'Azur, Géoazur, GRGS, Grasse, France

Definition

FTLRS. Acronym for French transportable laser ranging system (Nicolas et al., 2001).

ILRS. Acronym for International laser ranging service. International service established in 1998 (Pearlman et al., 2002). See entry *Geodesy, Networks and Reference Systems*.

Laser. Acronym for light amplification by stimulated emission of radiation.

Least-squares method. Optimization method. See (Tarantola, 2004).

LLR. Acronym for lunar laser ranging. See (Bender et al., 1973).

Introduction

The Satellite Laser Ranging (SLR) technique was born in 1964, with the first returns obtained from the satellite Beacon Explorer B. Since the launch of the first dedicated laser satellite Starlette in 1975 (Schutz et al., 1993; Barlier and Lefebvre, 2001), SLR has provided range measurements which have been the foundations of, not only precise orbit computations, but also major geodetic products, fundamental for many Geosciences applications. One of the advantages is that the onboard equipment is light, cheap, has an infinite lifetime, and does not consume any energy.

Technique

Principle and instrumentation

SLR is based on the measurement of the round-trip time of flight of laser pulses, between a ground station and a satellite orbiting round the Earth. As the measurements are carried out in the visible (or near visible) spectrum, the technique is dependent on weather conditions; it does not work under a cloudy sky. The satellite is equipped (if not entirely covered) with reflectors specially designed to reflect the laser in the incident direction. The ground station is composed of several instruments: a laser, a telescope for the emission and/or the reception of light, a detection system (for the start and the return of the laser), an event timer, a timing system for the chronometry, a frequency standard, a measurement calibration system (required because of the instrumental instabilities), a weather station (for the atmospheric correction, see Sect. Data analysis and orbitography), and a completely computerized system to pilot the telescope during the satellite tracking, to process the measurements, and to send the data obtained.

This complex ground instrumentation requires a permanent staff to (1) maintain the instruments and, consequently, to guarantee a stable quality of the measurements and to (2) track the satellites. This constant manpower requirement and the need for a cloudless sky to track satellites are drawbacks of the technique, even if some stations are now on the way toward automation. Moreover, these requirements make the SLR technique different (but not less effective, far from it) from the other satellite techniques: Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), Global

Positioning System (GPS) – see entry *GPS, Data Acquisition and Analysis*, GLObal NAvigation Satellite System (GLONASS), and the future Galileo (Europe) and Compass (China) systems. Indeed, these systems (will) operate whatever the weather conditions and their complexity lies more in the spatial segment.

Data

The raw observations produced during a satellite pass over a station are not directly put at the disposal of the scientific community. Indeed, with nearly 10 laser pulses per second, a significant number of returns can be obtained. Roughly speaking, these latter are thus compressed to deduce the so-called "normal points," gathering all the information contained in full-rate points, over specific spans of time, depending on the satellite altitude. The ILRS recommended algorithm is available at http://ilrs.gsfc.nasa.gov/products_formats_procedures/normal_point/np_algo.html. Normal point files per satellite can be retrieved through the ILRS data centers at ftp://cddis.gsfc.nasa.gov/slr/ and http://www.dgfi.badw-muenchen.de/edc/edc.html.

Technological evolution

Since the early seventies, the SLR technology and, consequently, the measurement quality have constantly improved, starting at a 1.5 m precision level, with a significant step toward the millimeter accuracy reached in the nineties (Degnan, 1993). Presently, a typical laser configuration is (1) a wavelength of 532 nm (green), (2) a pulse width ranging from 35 to 200 ps, and (3) a fire rate of 5, 10, or 20 Hz. The overall accuracy of the measurements is assessed at the 5 mm level for a few stations – see (Exertier et al., 2006) who also provide a complete error budget of the technique.

The principal technological improvements under study are the following (Noll and Pearlman, 2007, 6-1): reduction of the pulse length to 10 ps to improve the single shot precision; kHz laser ranging to improve the normal point precision (by providing a larger number of averaged raw data per normal point) and the speed of target acquisition during satellite tracking, together with more precise event timers (precision of 3 ps or less); multi-wavelength ranging to test possible improvements of the atmospheric correction.

Satellite constellation

Passive laser satellites have been developed for geodesy purposes. Their low area to mass ratios minimize the amplitude of the surface forces (atmospheric drag, for instance). The SLR network first tracks dedicated satellites, which have a practically infinite lifetime. The two first SLR satellites were Starlette, developed by CNES (France) in 1975, with a 812 km perigee height, and LAGEOS, launched by NASA (United States) in 1976, with a 5,860 km perigee height. They were followed by similar satellites in the eighties and in the nineties: Ajisai (1986, Japan, 1,490 km), Etalon-1 and -2 satellites (1989, USSR, 19,120 km), LAGEOS-2 (1992, United States and Italy, 5,620 km), and Stella (1993, France, 800 km). The constellation of SLR geodetic satellites is now composed of orbits with a wide range of altitudes and inclinations to the equator, providing various dynamical configurations. Some new SLR-only missions are under investigation, such as the LAGEOS III mission, for purposes concerning fundamental physics (Ciufolini, 1994).

The network also tracks (or has tracked) non-specific SLR satellites equipped with laser retroreflectors: altimetric (TOPEX/Poséidon, Jason-1, Jason-2), remote sensing (ERS-1 and -2, ENVISAT), gravimetric (CHAMP, GRACE-A and -B, GOCE), and GNSS satellites (some GPS, GLONASS, and Galileo – GIOVE-A and -B – satellites).

Currently, 34 artificial satellites are being tracked by the network. Moreover, few stations, equipped with large telescopes have the capability to measure the range to the Moon (LLR technique). One station is located in Europe (Grasse, France, telescope with a 1.6 m aperture), and two others are located in North America: at Apache point, in southern New Mexico, where the Apollo project utilizes a 3.5 m telescope, and at the McDonald Observatory, University of Texas (75 cm). See http://ilrs.gsfc.nasa.gov/satellite_missions/list_of_satellites/.

Ground network

Starting with five stations in the early seventies, the network configuration evolved over the last 35 years and the number of stations has been stable for several years. The current network is composed of about 30 stations (see http://ilrs.gsfc.nasa.gov/stations), distributed among 20 countries or so. Most of these stations observe only during night, a few having as well the capability of observing in daylight (useful for Global Navigation Satellite System – GNSS – satellite tracking, see Sect. Data analysis and orbitography). The distribution of this network, as is the case for the Very Long Baseline Interferometry (VLBI) network (see entry *Very Long Baseline Interferometry*), is generally admitted as a weakness of the technique. Indeed, in the northern hemisphere, most stations are located at mid-latitudes; in the southern hemisphere, there are only few stations (5, typically). Moreover, the densification and the maintenance of an ideal network are dependent on political decisions and the economic context (because of the financial commitments required to both build and maintain a SLR station).

This situation should improve in the near future, thanks to projects such as the NASA SLR2000 project (completely automated, eye-safe, highly reliable, and low-cost SLR stations), recently relaunched (McGarry and Zagwodski, 2008) and now called NGSLR, or thanks to mobile systems (FTLRS, for instance).

Data analysis and orbitography

General principle

A typical data analysis follows this two-step sketch, which is more or less the same for GPS and DORIS data analyses:

1. *Orbit computation, per satellite*. The physical forces acting on the satellite motion account for gravitational (Earth's gravity field, including the time-varying part, luni-solar attraction) and non-gravitational perturbations (mainly atmospheric drag and radiation pressure). The equations of motion are numerically propagated to provide time series of a state vector of orbital elements. The theoretical range rate is then computed, on the basis of these time series, and on geodetic products such as the Terrestrial Reference Frame (TRF), Earth Orientation Parameters – EOP, and some station displacement models. These "theoretical observations" are finally compared, in a least-squares sense, with the real data to estimate some dynamical parameters (initial satellite state vector, drag coefficients, etc.). Moreover, empirical accelerations can be estimated to compensate for some part of the modeling deficiencies (Crétaux et al., 1994) and, consequently, to improve the post-fit residual level. The computations are iterated until a given stop condition is fulfilled.
2. *Geodetic parameter estimation*. The orbital arcs are simultaneously used to derive predicted range measurements, together with a priori values of the parameters to be estimated and required models. Updates of the geodetic parameters (station positions, EOP, etc.), and generally of some of the dynamical parameters estimated during step 1, are then estimated in a least-square sense. If necessary, during this step, the measurements can be gathered for several different satellites and over a long period of time.

Most of the models applied (station displacements, Earth's gravity field, atmospheric density, etc.) are not specific to the SLR data processing. Their list is provided in the International Earth Rotation and Reference Systems Service (IERS) conventions (McCarthy and Petit, 2004).

Specific corrections

As they are carried out in the visible spectrum, SLR laser beams are not affected by the ionosphere crossing, in comparison to the radio-electric signals. But the tropospheric propagation still limits their accuracy. Recently, an improved model of zenith delay has been developed (Mendes and Pavlis, 2004). This model, together with an improved mapping function (Mendes et al., 2001), is currently the standard for SLR data analyses. This correction takes typical values of several meters and its inaccuracy is still assessed at a few-mm level (Exertier et al., 2006). In order to reach the 1-mm accuracy in the near future, improvements are under study; for example, ray tracing (Hulley and Pavlis, 2007) and multi-wavelength systems (Hamal et al., 2008).

The temporal spread of returned laser pulses due to the reflection from multiple reflectors on the satellite is another significant error at the 1-mm level. To better handle these effects, (Otsubo and Appleby, 2003) have recently proposed system-dependent center-of-mass corrections for spherical satellites (LAGEOS, Ajisai, and Etalon), with variations up to 10 mm. Efforts are still in progress in order to better understand and take these effects into account in SLR analyses (Luck et al., 2008, Sect. 8).

Despite the systematic in situ calibrations, ranges can be affected by systematic errors, dependent on the instrumentation: time and range biases. Due to their strong correlation with station heights, much attention is paid to range biases and their temporal variations. They are indeed constantly monitored in the framework of the ILRS activities – cf. (Otsubo et al., 2009) for instance. Moreover, analysis efforts towards better estimations of these biases are in progress, see (Appleby et al., 2009) as an example. Typical bias values range from a few millimeter to centimetric values. These biases can be handled in different ways (applied and/or estimated) during any data processing.

Finally, in a relativistic context, the propagation of light between a ground station and a near-Earth satellite induces a supplementary time delay with respect to the purely geometric path (Ries et al., 1988). This correction takes few-mm values.

LAGEOS satellites

Due to their relatively high altitudes (about 6, 000 km), both LAGEOS satellites are essential targets and their orbits are permanently tracked and computed with great care. Since their launch, the quality of the computed orbits has evolved with the technique: from a 50 cm precision in 1976 (Exertier et al., 2006) to a few centimeters in the mid nineties (Marshall et al., 1995) to nearly 1 cm nowadays. Additionally, various non-gravitational effects have been evidenced in the LAGEOS-1 orbit: solar radiation pressure, Yarkovsky–Shah thermal effect, asymmetric reflectivity of the satellite surface, and asymmetric thermal emissivity of the Earth (Métris et al., 1999).

GNSS satellites

Since the first study (Degnan and Pavlis, 1994), and sometimes through dedicated campaigns such as IGEX-98 (Barlier et al., 2001), SLR has always contributed to the validation of GNSS orbits. This validation generally consists in the comparison between SLR range measurements to GNSS satellites and the predicted ranges computed with the GNSS-only orbits. According to (Urschl et al., 2007), the consistency (assessed over nearly four years) between SLR and GPS (resp. GLONASS) is at an accuracy level of −3 cm (resp. −3 mm) with a precision at the level of 2.5 cm (resp. 5 cm). Further studies are in progress to get an exhaustive understanding of the inter-technique differences. Furthermore, all satellites of the future GNSS

Galileo system with be equipped with Laser Ranging Arrays (LRA) for such orbit validations.

Applications to geodesy and geodynamics

Earth's gravity field

Before the launch of dedicated space missions (CHAMP, GRACE, and GOCE), the gravity field models were mainly built on the basis of SLR data acquired on the whole constellation, providing various dynamical configurations (semi-major axes, inclinations to the equator), and, consequently, various sensitivities to the mass distribution within the Earth's system. The GRIM5-S1 model (Biancale et al., 2000), one of the latest SLR-only based, was prepared in a joint German–French effort. The solution was obtained by solving simultaneously for the gravitational and ocean tide potentials and tracking station positions. Nowadays, long wavelengths of the gravity field are still deduced from SLR data, accurate enough to provide a value of the Earth's gravitational constant with 10 significant digits (corresponding to an orbital post-fit residual error of the order of a few centimeters), see (Tapley et al., 1993). Since the CHAMP mission was launched in 2000, followed by the GRACE mission in 2002, and GOCE in 2009, a new generation of nearly monthly Earth's gravity field models was initiated, benefiting both from the high sensitivity of the orbit (a few hundred km high) to regional variations and from accelerometric data, measuring the non-gravitational forces (mainly atmospheric drag). The GRACE-only models expanded in spherical harmonics provide coefficients up to degree and order 360. See (Reigber et al., 2003; Lemoine et al., 2007; Tapley et al., 2005) and entries *Gravity Field of the Earth* and *Gravity Field, Temporal Variations from Space Techniques*.

Mass variations within the Earth's dynamic system have a temporal spectrum ranging from hours to decades, and even longer, that can be evidenced by SLR data, because they are available over a very long period of time. Many of them are related to both long-term and short-term climate forcing, and are linked to mass redistribution in the atmosphere, ocean, and continental water (Chen et al., 2000; Cheng and Tapley, 2004). By the next years, SLR data, moreover, are likely to provide an unique opportunity to continue to monitor the Earth's gravity field, after the end-of-life of the CHAMP, GRACE, and GOCE missions.

Terrestrial reference frame and Earth's rotation

Since the first computation of a combined TRF in 1984 (called BTS84), SLR has always been a keystone of the BTS (BIH – Bureau International de l'Heure – Terrestrial System) and, after 1988, of the International Terrestrial Reference System (ITRS, see entry *Geodesy, Networks and Reference Systems*) realizations (McCarthy and Petit, 2004, Chap. 4). Indeed, their origin and scale definitions have always strongly involved this technique. Parallel to the SLR-positioning precision, starting with a decimetric precision (BTS84), the successive ITRF versions have turned 1 cm-precision level with the ITRF96. For the more recently published version (ITRF2005), the whole precision of positions (resp. velocities) is under 2 mm (resp. 1 mm/y). And, for the first time of the ITRF history, the ITRF2005 was computed with times series of (weekly for SLR) station positions and (daily for SLR) EOP. Regarding positioning, the weekly WRMS for the well performing SLR stations is 5 mm for both planimetry and height (Altamimi et al., 2007). For comparison, the weekly WRMS for GPS are 2 mm (resp. 5 mm) for planimetry (resp. for height). It must be noted that the most recent ITRS realization (ITRF2008) is now available (http://itrf.ign.fr/ITRF_solutions/2008).

Regarding Earth's rotation, SLR provides (with the LAGEOS data) the longest space-geodetic time series of EOP (Gross, 2009). Starting with a 10 milliarcsecond (mas) precision in the seventies, the discrepancy between the SLR and the IERS polar motion series was assessed at the level of 0.3 mas in the mid 2000 years (Gambis, 2004). (Coulot et al., 2010) have recently showed that significant precision improvements are still possible with the help of a rigorous referencing.

Global Earth's deformations

For many years, SLR has provided unique and inestimable information about the geocenter motion (Pavlis and Kuźmicz-Cieślak, 2009a). The geocenter motion is the 3-dimensional motion of the instantaneous Earth's center-of-mass with respect to any conventional reference (the ITRF secular origin, for instance). It is related to the responses of the Earth to the displacements of its fluid masses (atmosphere, ocean, etc.) and its temporal variations are dominated by annual signals of few mm amplitudes. (Collilieux et al., 2009) provide an exhaustive review about this geodynamical signal and, moreover, assess the effect of the SLR network on its determination.

The ITR2005 computation has been a good opportunity to study the residual position time series which contain the non-linear part of the station motions. Regarding SLR station positions, most of the time series evidence annual signals. As surface loading effects (see van Dam et al. (2002) and http://www.sbl.statkart.no/) are not currently modeled in the routine data processing, they are supposed to be evidenced by the position series. (Collilieux et al., 2010) show that an a posteriori correction for these effects with a geodynamical model leads to a reduction of the amplitudes of the annual signals for a majority of stations. Furthermore, (Pavlis et al., 2009) strongly recommend the modeling of such effects in the future SLR data processing to achieve a better accuracy.

Altimetry and sea level monitoring

Stable and accurate TRF, satellites orbits, and altimeter measurements are prerequisites for robust altimetric Mean Sea Level (MSL) computations over large time periods

(Leuliette et al., 2004; Beckley et al., 2007; Ablain et al., 2009). Indeed, these conditions are crucial to connect the data provided by successive missions. Since the advent of altimetry (launch of TOPEX/Poséidon in 1992), SLR has contributed to all these efforts, first regarding the TRF (see Sect. Terrestrial reference frame and Earth's rotation). SLR has also played a major role in the Precision Orbit Determination (POD) framework, additionally to DORIS and GPS data. For instance, the technique has recently contributed to the achievement of a 1-cm radial accuracy for the orbit of Jason-1 (Luthcke et al., 2003). In this context, the laser-based short-arc technique (Bonnefond et al., 1999) is also a major method to assess the radial accuracy of the altimetric satellites on an operational basis. Finally, thanks to thoroughly instrumented sites, it is possible to monitor both bias and drift of the radar altimeters, by taking advantage of direct overflight passes (see http://www.gavdos.tuc.gr/intro.jsp for an illustration of the calibration principle). The French campaigns led in Corsica with the FTLRS show that the quality of such a calibration mainly depends on the accurate positioning (mainly deduced from Starlette and Stella data) of the mobile system and on the stability of its range bias (Gourine et al., 2008).

Summary

Observing and understanding the system Earth require numerous and accurate geodetic measurements over both spatial and temporal wide spectra. Since the early seventies, the SLR technique has provided such measurements, inestimable for the progress of Space Geodesy and, consequently, of Earth's sciences. Nowadays, in the Global Geodetic Observing System (GGOS) context, and, more particularly, with the stringent requirements on the ITRF accuracy (1 mm and 0.1 mm/y) and the crucial need of taking advantage of the features of each space-geodetic technique, SLR still has a major role to play.

Moreover, to reach the ambitious goals fixed for the near future, this technique is currently renewed. From a technological point of view, the kHz ranging has recently showed the wide spectrum of its abilities, see the dedicated sections in (Luck et al., 2008; Schillak, 2009). Regarding the ground and spatial segments, the design of the ideal SLR network of the future is in progress (Pavlis and Kuźmicz-Cieślak, 2009b), as is the design of the next generation of SLR stations, and new satellite missions are already planned. Concerning the data analysis, new modelings should help to improve the accuracy of the products (Pavlis et al., 2009). More particularly, the unprecedented quality of the Earth's gravity field models reached with the GRACE (Tapley et al., 2004) and GOCE missions should allow the use of low-orbiting satellite (Starlette, Stella, and Ajisai) data for station position computations (Lejba and Schillak, 2009).

Finally, new fundamental objectives have recently been fixed for the SLR technique in the time (frequency) domain, though time transfer projects (Samain et al., 2009) and interplanetary navigation (Ni et al., 2002).

Bibliography

Ablain, M., Cazenave, A., Valladeau, G., and Guinehut, S., 2009. A new assessment of the error budget of global mean sea level rate estimated by satellite altimetry over 1993–2008. *Ocean Science*, **5**, 193–201.

Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B., and Boucher, C., 2007. ITRF2005: a new release of the international terrestrial reference frame based on time series of station positions and earth orientation parameters. *Journal of Geophysical Research*, **112**, B09401.

Appleby, G., Wilkinson, M., Luceri, V., Gibbs, P., and Smith, V., 2009. Attempts to separate apparent observational range bias from true geodetic signals. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Barlier, F., Berger, C., Bonnefond, P., Exertier, P., Laurain, O., Mangin, J. F., and Torre, J. M., 2001. Laser-based validation of GLONASS orbits by short-arc technique. *Journal of Geodesy, Special issue IGEX-98*, **11**, 600–612.

Barlier, F., and Lefebvre, M., 2001. A new look at planet Earth: satellite geodesy and geosciences. In *The Century of Space Science*. The Netherlands: Kluwer, pp. 1623–1651.

Beckley, B. D., Lemoine, F. G., Luthcke, S. B., Ray, R. D., and Zelensky, N. P., 2007. A reassessment of global and regional mean sea level trends from TOPEX and Jason-1 altimetry based on revised reference frame and orbits. *Geophysical Research Letters*, **34**, L14608.

Bender, P. L., Currie, D. G., Dicke, R. H., Eckhardt, D. H., Faller, J. E., Kaula, W. M., Mulholland, J. D., Plotkin, H. H., Poultney, S. K., Silverberg, E. C., Wilkinson, D. T., Williams, J. G., and Alley, C. O., 1973. The lunar laser ranging experiment. *Science*, **182**, 229–238.

Biancale, R., Balmino, G., Lemoine, J. M., Marty, J. C., Moynot, B., Barlier, F., Exertier, P., Laurain, O., Gegout, P., Schwintzer, P., Reigber, C., Bode, A., Gruber, T., König, R., Massmann, F. H., Raimondo, J. C., Schmidt, R., and Zhu, S. Y., 2000. A new global earth's gravity field model from satellite orbit perturbations: GRIM5-S1. *Geophysical Research Letters*, **27**, 3611–3614.

Bonnefond, P., Exertier, P., and Barlier, F., 1999. Geographically correlated errors observed from a laser-based short-arc technique. *Journal of Geophysical Research*, **104**(C7), 15885–15893.

Chen, J. L., Wilson, C. R., Eanes, R. J., and Tapley, B. D., 2000. A new assessment of long-wavelength gravitational variations. *Journal of Geophysical Research*, **105**(B7), 16271–16277.

Cheng, M. K., and Tapley, B. D., 2004. Variations in the earth's oblateness during the past 28 years. *Journal of Geophysical Research*, **109**(B9), B09402.

Ciufolini, I., 1994. Gravitomagnetism and status of the LAGEOS III experiment. *Classical and Quantum Gravity*, **11**(6A), 73–81.

Collilieux, X., Altamimi, Z., Ray, J., van Dam, T., and Wu, X., 2009. Effect of the satellite laser ranging network distribution on geocenter. *Journal of Geophysical Research*, **114**, B04402.

Collilieux, X., Altamimi, Z., Coulot, D., van Dam, T., and Ray, J., 2010. Impact of loading effects on determination of the international terrestrial reference frame. *Advances in Space Research*, **45**, 144–154.

Coulot, D., Pollet, A., Collilieux, X., and Berio, P., 2010. Global optimization of core station networks for space geodesy: application to the referencing of the SLR EOP with respect to ITRF. *Journal of Geodesy*, **84**, 31–50.

Crétaux, J. F., Nouel, F., Valorge, C., and Janniere, P., 1994. Introduction of empirical parameters deduced from the Hill's equations for satellite orbit determination. *Manuscripta Geodaetica*, **19**, 135–156.

Degnan, J. J., 1993. Millimeter accuracy satellite laser ranging: a review. In: *Contributions of Space Geodesy to Geodynamics: Technology, Geodynamic Series*. Washington, D. C.: AGU, Vol. 25, pp. 133–162.

Degnan, J. J., and Pavlis, E. C., 1994. Laser ranging to GPS satellites with centimeter accuracy. *GPS World*, **5**(9), 62–70.

Exertier, P., Bonnefond, P., Deleflie, F., Barlier, F., Kasser, M., Biancale, R., and Ménard, Y., 2006. Contribution of laser ranging to earth's sciences. *Comptes Rendus Geoscience*, **338**, 958–967.

Gambis, D., 2004. Monitoring earth orientation using space geodetic techniques: state-of-the-art and prospective. *Journal of Geodesy*, **78**(4–5), 295–303.

Gourine, B., Kahlouche, S., Exertier, P., Berio, P., Coulot, D., and Bonnefond, P., 2008. Corsica SLR positioning campaigns (2002 and 2005) for satellite altimeter calibration missions. *Marine Geodesy*, **31**, 103–116.

Gross, R. S., 2009. Laser ranging contributions to earth rotation studies. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Hamal, K., Prochazka, I., Blazef, J., Fumin, Y., Jingfu, H., Zhongping, Z., Kunimori, H., Greene, B., Kirchner, G., Koidl, F., Riepfel, S., and Gurtner, W., 2008. Multi color satellite laser ranging at Czech technical university. In Luck et al. (eds.), *Proceedings of the 15th International Workshop on Laser Ranging*, Canberra, Australia, October 16–20, 2006, EOS Space Systems Pty. Ltd.

Hulley, G. C., and Pavlis, E. C., 2007. A ray-tracing technique for improving satellite laser ranging atmospheric delay corrections, including the effects of horizontal refractivity gradients. *Journal of Geophysical Research*, **112**, B06417.

Lejba, P., and Schillak, S., 2009. Determination of the SLR station coordinates and velocities on the basis of laser observations of low satellites. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Lemoine, J. M., Bruinsma, S., Loyer, S., Biancale, R., Marty, J. C., Perosanz, F., and Balmino, G., 2007. Temporal gravity field models inferred from GRACE data. *Advances in Space Research*, **39**(10), 1620–1629.

Leuliette, E. W., Nerem, R. S., and Mitchum, G. T., 2004. Calibration of TOPEX/Poséidon and Jason altimeter data to construct a continuous record of mean sea level change. *Marine Geodesy*, **27**(1), 79–94.

Luck, J., Moore, C., and Wilson, P. (eds.), 2008. *Proceedings of the 15th International Workshop on Laser Ranging*, Canberra, Australia, October 16–20, 2006, EOS Space Systems Pty. Ltd.

Luthcke, S. B., Zelensky, N. P., Rowlands, D. D., Lemoine, F. G., and Williams, T. A., 2003. The 1-centimeter orbit: Jason-1 precision orbit determination using GPS, SLR, DORIS and altimeter data. *Marine Geodesy*, **26**, 399–421.

Marshall, J. A., Klosko, S. M., and Ries, J. C., 1995. Dynamics of SLR tracked satellites. *Reviews of Geophysics, Supplement*, U.S. national report to IUGG 1991–1994, pp. 353–360.

McCarthy, D. D., and Petit, G. (eds.), 2004. IERS conventions (2003). IERS Technical Note 32.

McGarry, J., and Zagwodski, T., 2008. SLR2000: the path toward completion. In Luck et al. (eds.), *Proceedings of the 15th International Workshop on Laser Ranging*, Canberra, Australia, October 16–20, 2006, EOS Space Systems Pty. Ltd.

Mendes, V. B., Prates, G., Pavlis, E. C., Pavlis, D. E., and Langley, R. B., 2001. Improved mapping functions for atmospheric refraction correction in SLR. *Geophysical Research Letters*, **29**(10), 1414.

Mendes, V. B., and Pavlis, E. C., 2004. High-accuracy zenith delay prediction at optical wavelengths. *Geophysical Research Letters*, **31**, L14602.

Métris, G., Vokroulicky, D., Ries, J. C., and Eanes, R. J., 1999. LAGEOS Spin axis and non-gravitational excitations of its orbit. *Advances in Space Research*, **23**(4), 721–725.

Ni, W. T., Zhu, J., Wu, X. P., Chu, G. B., Yang, B., Gao, J., Guan, M., Tang, C. J., Chou, Y., Chang, C. H., Huang, T., Qu, Q. Y., Yi, Z. H., Li, G., Tao, J., Wu, A. M., Luo, J., Yeh, H. C., Zhou, Z. B., Xiong, Y., Bi, S. L., Xu, C., Wu, X. J., Tang, M. X., Bao, Y., Li, F., Huang, C., Yang, F. M., Ye, S. H., Zhang, S. L., Zhang, Y. Z., Nie, Y., Chen, G., Christensen-Dalsgaard, J., Dittus, H., Fujii, Y., Lämmerzahl, C., Mangin, J. F., Peters, A., Rüdiger, A., Samain, E., and Schiller, S., 2002. Mini-ASTROD: mission concept. *International Journal of Modern Physics D*, **11**, 1035–1348.

Nicolas, J., Pierron, F., Samain, E., and Barlier, F., 2001. Centimeter accuracy for the French transportable laser ranging station (FTLRS) through sub-system controls. *Surveys in Geophysics*, **22**, 449–464.

Noll, C., and Pearlman, M. (eds.), 2007. *International Laser Ranging Service 2005–2006 report*. NASA/TP-2007-214153.

Otsubo, T., and Appleby, G. M., 2003. System-dependent center-of-mass correction for spherical geodetic satellites. *Journal of Geophysical Research*, **108**(B4), 2201.

Otsubo, T., Kobayashi, M., Gotoh, T., and Kubo-oka, T., 2009. Multi-satellite daily bias report: how to read and handle it. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Pavlis, E. C., and Kuźmicz-Cieślak, M., 2009a. Geocenter motion: causes and modeling approaches. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Pavlis, E. C., and Kuźmicz-Cieślak, M., 2009b. SLR and the next generation global geodetic networks. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Pavlis, E. C., Kuźmicz-Cieślak, M., and Ninkey, P. M., 2009. Improved modeling approaches towards the mm SLR. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Pearlman, M. R., Degnan, J. J., and Bosworth, J. M., 2002. The international laser ranging service. *Advances in Space Research*, **30**(2), 135–143.

Reigber, C., Schwintzer, P., Neumayer, H., Barthelmes, F., König, R., Förste, C., Balmino, G., Biancale, R., Lemoine, J. M., Loyer, S., Bruinsma, S., Perosanz, F., and Fayard, T., 2003. The CHAMP-only earth gravity field model EIGEN-2. *Advances in Space Research*, **31**(8), 1833–1888.

Ries, J. C., Huang, C., and Watkins, M. M., 1988. The effect of general relativity on near-earth satellites in the solar system Barycentric and Geocentric reference frames. *Physical Review Letters*, **61**, 903–906.

Samain, E., Guillemot, P., Exertier, P., Albanese, D., Berio, P., Laurain, O., Para, F., Paris, J., Torre, J. M., Viot, H., Vrancken, P., Petitbon, I., and Leon, S., 2009. Time transfer by laser link – T2L2: first data. In Schillak, S. (ed.), *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.

Schillak, S. (ed.), 2009. *Proceedings of the 16th International Workshop on Laser Ranging*, Poznán, Poland, October 13–17, 2008, Space Research Centre, Polish Academy of Sciences.
Schutz, B. E., Cheng, M. K., Eanes, R. J., Shum, C. K., and Tapley, B. D., 1993. Geodynamic results from starlette orbit analysis. In Smith, D. E., and Turcotte, D. L. (eds.), *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*. Geodynamics Series, Vol. 24, pp. 175–190.
Tapley, B. D., Schutz, B. E., Eanes, R. J., Ries, J. C., and Watkins, M. M., 1993. Lageos laser ranging contributions to geodynamics, geodesy, and orbital dynamics. In Smith, D. E., and Turcotte, D. L. (eds.), *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*. Geodynamics Series, Vol. 24, pp. 147–174.
Tapley, B. D., Bettadpur, S., Watkins, M., and Reigber, C., 2004. The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, **31**(9), L09607.
Tapley, B., Ries, J., Bettadpur, S., Chambers, D., Cheng, M., Condi, F., Gunter, B., Kang, Z., Nagel, P., Pastor, R., Pekker, T., Poole, S., and Wang, F., 2005. GGM02 – An improved earth gravity field model from GRACE. *Journal of Geodesy*, **79**(8), 467–478.
Tarantola, A., 2004. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: SIAM.
Urschl, C., Beutler, G., Gurtner, W., Hugentobler, U., and Schaer, S., 2007. Contribution of SLR tracking data to GNSS orbit determination. *Advances in Space Research*, **39**, 1515–1523.
van Dam, T., Plag, H. P., Francis, O., and Gegout, P., 2002. GGFC special bureau for loading: current status and plans. *IERS Technical Note*, **20**, 180–198.

Cross-references

Earth Rotation
Geodesy, Figure of the Earth
Geodesy, Networks and Reference Systems
Geodesy, Physical
GPS, Data Acquisition and Analysis
Gravity Field of the Earth
Gravity Field, Temporal Variations from Space Techniques
Gravity Field, Time Variations from Surface Measurements
Gravity, Global Models
Spherical Harmonic Analysis Applied to Potential Fields
Very Long Baseline Interferometry

SEAFLOOR SPREADING

Richard N. Hey
Hawaii Institute of Geophysics and Planetology, School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, HI, USA

Definition

Seafloor spreading is the mechanism by which new oceanic lithosphere is created at and moves away from divergent plate boundaries. The seafloor spreading hypothesis led to one of the most important paradigm shifts in the history of the Earth sciences.

Introduction

The revolutionary seafloor spreading hypothesis improved and subsumed the continental drift hypothesis, and rapidly culminated in what is now known as the plate tectonics scientific revolution. It was first hypothesized by Hess in 1960, in a widely circulated preprint and paper (Hess, 1962) he considered so speculative that he called it "an essay in geopoetry". This hypothesis was named "seafloor spreading" in another influential early paper (Dietz, 1961). It offered a simple explanation for many problems with the prevailing paradigm that the Earth was a mostly static, slowly contracting planet, with fixed continents and old ocean basins, and no large-scale horizontal displacements. This paradigm had previously been challenged, most notably by Wegener's continental drift hypothesis (Wegener, 1912), and by paleomagnetic measurements in the 1950s that were consistent with continental drift, but before the 1960s these ideas were not generally accepted.

The revolution

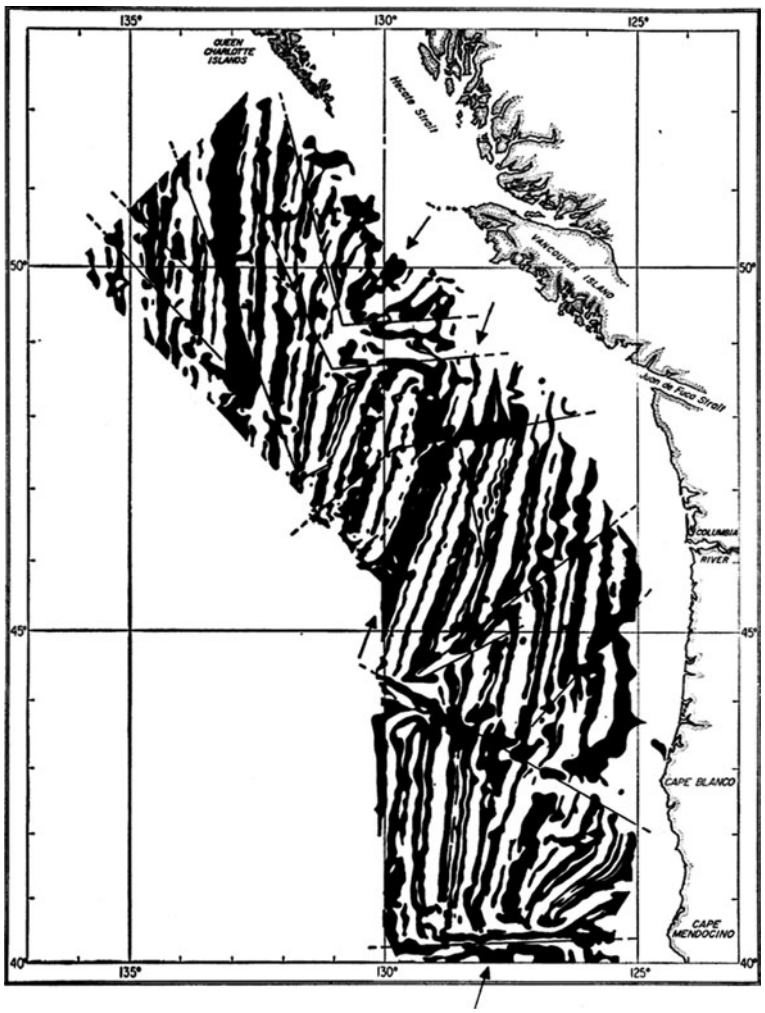
Hess realized that if mantle convection carried seafloor and continents away from seafloor spreading centers (mid-ocean ridges) toward trenches (subduction zones), with new seafloor formed at ridge axes and destroyed at trenches, this would explain the shallow bathymetry, earthquakes, high heat flow, lack of sediments, and extensional structures characterizing ridge axes, as well as the deep trenches, earthquakes, compressional structures, mountain ranges, and volcanoes characterizing subduction zones.

A key step in the confirmation of seafloor spreading was the recognition by Vine and Matthews (1963) (and independently by Morley, in a paper unfortunately turned down as too speculative by both *Nature* and the *Journal of Geophysical Research*, eventually published in Morley and Larochelle, 1964), which offered a simple explanation for the existence of puzzling magnetic "stripes" on the seafloor. The Vine–Matthews (or Vine–Matthews–Morley) hypothesis proposed that a combination of seafloor spreading and episodic reversals of the Earth's magnetic field (at the time another very speculative idea) would create alternating zones of normally and reversely magnetized crust, and thus linear positive and negative magnetic anomalies in a pattern symmetric about the spreading axis.

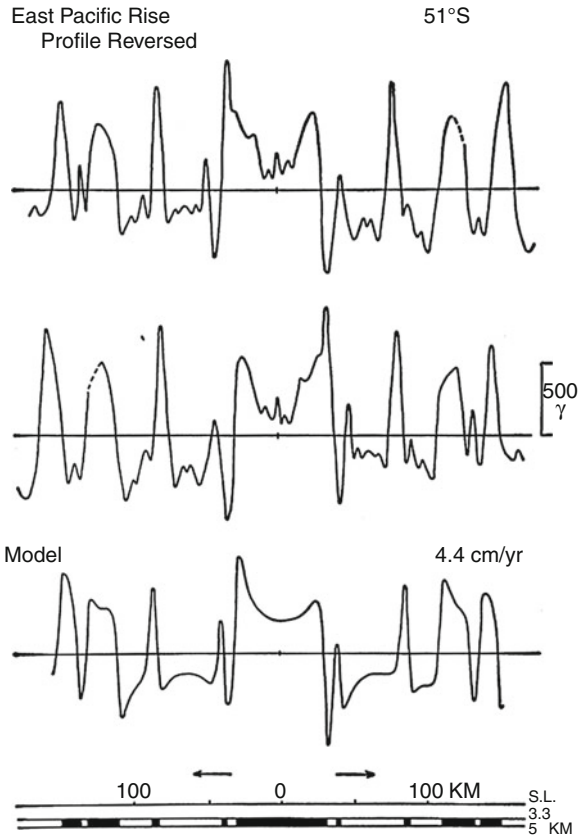
The next key step in the revolution occurred in 1965, when Wilson noted that deformation of the Earth's crust is concentrated in narrow mobile belts, and postulated that these features are all interconnected in a global network, the first qualitative model of plate tectonics (Wilson, 1965). The zones of extension and compression are connected by a new class of faults defined by Wilson as transform faults, which required relative plate motion and turned out to be the most important type of fault on Earth. He showed that seafloor spreading occurring on offset mid-ocean ridge axes would produce relative motion exactly opposite to the motion the sense of offset would predict without seafloor spreading, and that earthquakes should only occur between the offset seafloor spreading axes. These radical predictions, completely opposite to prevailing wisdom, were soon confirmed seismically. Furthermore, by correctly interpreting the San Andreas fault

as a transform fault between the Pacific and North American plates, he predicted the existence of previously unrecognized seafloor spreading offshore western North America. Vine and Wilson (1965) showed that the predicted magnetic symmetry existed in this area, where magnetic stripes had been discovered in the mid-1950s (Mason, 1958; Mason and Raff, 1961; Raff and Mason, 1961), over what they recognized as the Juan de Fuca Ridge (Figure 1), and furthermore, that the pattern of stripes corresponded perfectly with the pattern of magnetic field reversals (Cox et al., 1963; McDougall and Tarling, 1963), once the Jaramillo anomaly was discovered. Vine (1966) also demonstrated similar symmetry and correlation with the reversal timescale in another important data set, the Project Magnet aeromagnetic data collected over the Reykjanes Ridge south of Iceland. His color figures of these classic data sets and correlations (Vine, 1968) became iconic.

The smoking gun for many scientists was the discovery by Pitman and Heirtzler (1966) of near perfect symmetry in the Eltanin-19 profile collected over the Pacific–Antarctic Ridge. Essentially every tiny wiggle seen in the magnetic anomaly profile on the Pacific plate was mirrored on the Antarctic plate, and correlated perfectly with the magnetic reversal time scale (Figure 2). The symmetry in these data required a symmetric axial process, with new seafloor carried away on both plates, and thus provided compelling evidence for both the Vine–Matthews and seafloor spreading hypotheses. Vine convincingly summarized this evidence in influential symposia and publications (Vine, 1966, 1968), and, by the end of 1966, seafloor spreading was generally accepted by marine geophysicists, who quickly extrapolated the magnetic reversal time scale from ~10 Ma to ~80 Ma and worked out at least the basic recent evolutionary history of every ocean basin.



Seafloor Spreading, Figure 1 Raff and Mason (1961) magnetic stripes recognized as symmetric about seafloor spreading axes (*arrows*) by Vine and Wilson (1965). (From Vine, 1966. Reprinted with permission from AAAS.)



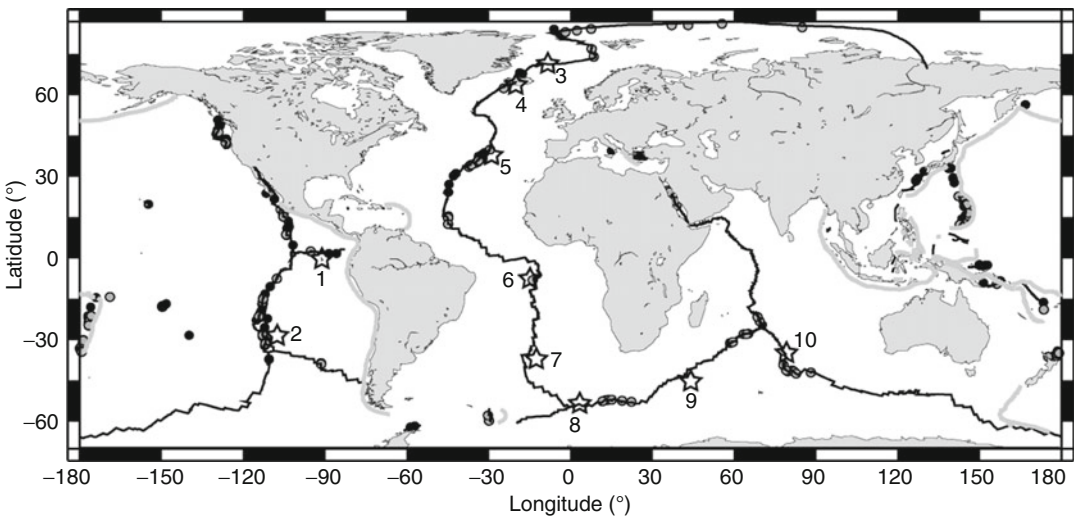
Seafloor Spreading, Figure 2 Eltanin-19 magnetic anomaly profile (*top*) from Pitman and Heirtzler (1966), shown compared with its mirror image (*middle*) and magnetic anomaly model calculated from the magnetic reversal timescale (*bottom*), assuming seafloor spreading at a constant 44 km/Myr. (From Vine, 1966. Reprinted with permission from AAAS.)

This scientific revolution culminated in plate tectonics the following year when Morgan (1968) and McKenzie and Parker (1967) made the key assumption that plates behave rigidly, and extended Wilson's transform fault geometry to the sphere. They showed that transforms are small circles about rotation poles describing relative plate motion, that seafloor spreading rates increase as the sine of the angular distance away from these poles, and that it was possible to use known patterns of seafloor spreading to quantitatively predict other plate motions and plate boundary evolution.

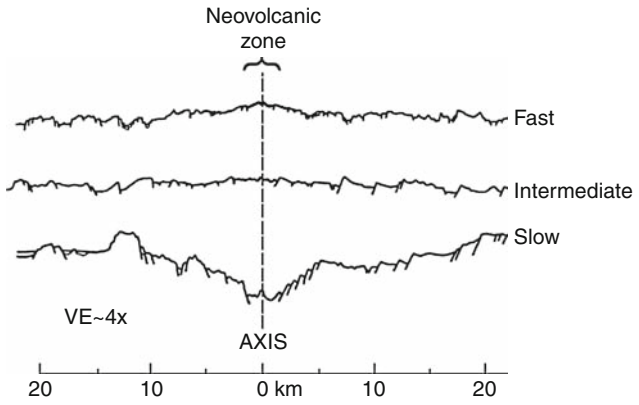
Ridge axis geometry, morphology, and crustal formation

Volcanism along the mid-ocean ridge system has formed the longest mountain chain on Earth (Figure 3). As the plates slowly (~0–150 mm/yr, or km/Myr, at present, DeMets et al., 2010, with rates remarkably similar whether using magnetic anomalies or geodetic measurements such as GPS) move apart by seafloor spreading, magma from the hot asthenosphere rises to fill the crack at mid-ocean ridge axes, elevated because of the higher temperatures. Both axial depth and morphology correlate with spreading rate (Macdonald, 1982). Fast-spreading ridge axes such as the East Pacific Rise generally have shallow depths and relatively smooth morphologies, with very small extensional axial summit troughs (except where these are completely buried by the latest eruption in the neovolcanic rift zone). Slow spreading ridges such as the Mid-Atlantic Ridge generally have much deeper, rougher, and higher-amplitude axial valleys (Figure 4), except near hotspots such as Iceland where magma supply is unusually large and even slow-spreading ridges have shallow axes with fast-spreading morphology.

The asthenosphere typically melts to become mid-ocean ridge basalt in a magma chamber under the ridge



Seafloor Spreading, Figure 3 Global distribution of seafloor spreading axes (*heavy black lines*) with known (*black dots*) or inferred (*gray dots*) hydrothermal fields. Stars are near-ridge hotspots. From NOAA/PMEL Vents Program website: http://www.pmel.noaa.gov/pubs/outstand/bake2544/images/fig01a.gif (Baker and German, 2004).



Seafloor Spreading, Figure 4 Ridge axis fault patterns and morphology as a function of spreading rate. (Modified from Macdonald, 1982.)

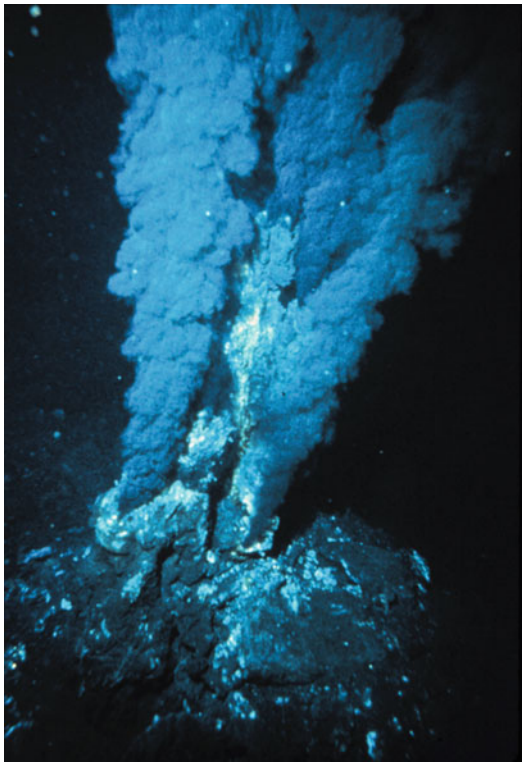
axis. Some basalt is erupted onto the seafloor to form pillow basalts and lava flows, which are progressively covered by sediments as the lithosphere ages and moves away from the axis. Below the basalts are the sheeted dikes, the paths through which the lava moved from magma chambers to the surface, and below the dikes are the layered gabbros which cooled and crystallized in place (intrusives) rather than erupting at axial volcanoes (extrusives). This typical sequence of mantle ultramafics – gabbro – sheeted dikes – extrusive basalts – sediments is observed on the seafloor in tectonic windows where existing lithosphere is rifted apart, and also where slices of old ocean lithosphere are thrust into ophiolite mountain belts as in Oman and Cyprus (Karson, 2002).

Hydrothermal vents

As the new seafloor cools it contracts and cracks. Water goes down these cracks, is heated, reacts with the surrounding rocks, and comes back up as buoyant hydrothermal vents (Baker and German, 2004). These occur at a range of chemistries and temperatures, up to ~400°C in black smokers, so-called because of the sulfide-rich plumes that precipitate suddenly when the superheated water is injected into the surrounding ~2°C seawater, forming sulfide chimneys (Figure 5). White and clear smokers occur at lower temperatures. The chemical reactions provide energy for an unusual kind of life that does not depend on photosynthesis but flourishes as chemosynthetic communities at many of these vents, rare deep-ocean oases in an otherwise extremely barren environment (Kelley et al., 2002).

Summary

Seafloor spreading was a critical step in the contentious scientific revolution from the previous static Earth paradigm to the now universally accepted plate tectonic paradigm. Today it refers to the processes creating new oceanic lithosphere where plates move apart. Seafloor spreading replaces the lithosphere destroyed by subduction, and exerts important influences on Earth's chemical and biological evolution.



Seafloor Spreading, Figure 5 Black smoker hydrothermal vents and sulfide chimneys. (From NOAA website: http://www.photolib.noaa.gov/htmls/nur04506.htm.)

Bibliography

Baker, E. T., and German, C. R., 2004. On the global distribution of hydrothermal vent fields. In German, C. R., et al. (eds.), *Mid-Ocean Ridges: Hydrothermal Interactions Between the Lithosphere and Oceans*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 148, pp. 245–266.

Cox, A., Doell, R. R., and Dalrymple, G. B., 1963. Geomagnetic polarity epochs and Pleistocene geochronometry. *Nature*, **198**, 1049–1051.

DeMets, C., Gordon, R. G., and Argus, D. F., 2010. Geologically current plate motions. *Geophysical Journal International*, **181**(1), 1–80, doi:10.1111/j.1365-246X.2009.04491.x.

Dietz, R. S., 1961. Continent and ocean basin evolution by spreading of the sea floor. *Nature*, **190**, 854–857.

Hess, H. H., 1962. History of ocean basins. In Engel, A. E. J., James, H. L., and Leonard, B. F. (eds.), *Petrologic Studies: A Volume to Honor A.F. Buddington*. New York: Geological Society of America, pp. 599–620.

Karson, J. A., 2002. Geologic structure of the uppermost oceanic crust created at fast- to intermediate-rate spreading centers. *Annual Review of Earth and Planetary Sciences*, **30**, 347–384.

Kelley, D. S., Baross, J. A., and Delaney, J. R., 2002. Volcanoes, fluids and life at mid-ocean ridge spreading centers. *Annual Review of Earth and Planetary Sciences*, **30**, 385–491.

Macdonald, K. C., 1982. Mid-ocean ridges: fine scale tectonic, volcanic and hydrothermal processes within the plate boundary zone. *Annual Review of Earth and Planetary Sciences*, **10**, 155–190.
Mason, R. G., 1958. A magnetic survey over the west coast of the United States between latitudes 32° and 36° N, longitudes 121° and 128° W. *Geophysical Journal of the Royal Astronomical Society*, **1**, 320–329.
Mason, R. G., and Raff, A. D., 1961. A magnetic survey off the west coast of North America 32° N to 42° N. *Bulletin of the Geological Society of America*, **72**, 1259–1265.
McDougall, I., and Tarling, D. H., 1963. Dating of polarity zones in the Hawaiian Islands. *Nature*, **200**, 54–56.
McKenzie, D. P., and Parker, R. L., 1967. The North Pacific: an example of tectonics on a sphere. *Nature*, **216**, 1276–1280.
Morgan, W. J., 1968. Rises, trenches, great faults and crustal blocks. *Journal of Geophysical Research*, **73**, 1959–1982.
Morley, L. W., and Larochelle, A., 1964. Paleomagnetism as a means of dating geological events. In *Geochronology in Canada*. Royal Society of Canada Special Publication, Vol. 8, pp. 39–50.
Pitman, W. C., III, and Heirtzler, J. R., 1966. Magnetic anomalies over the Pacific-Antarctic Ridge. *Science*, **154**, 1164–1171.
Raff, A. D., and Mason, R. G., 1961. Magnetic survey off the west coast of the United States between 40°N latitude and 52°N latitude. *Bulletin of the Geological Society of America*, **72**, 1267–1270.
Vine, F. J., 1966. Spreading of the ocean floor: new evidence. *Science*, **154**, 1405–1415.
Vine, F. J., 1968. Magnetic anomalies associated with mid-ocean ridges. In Phinney, R. A. (ed.), *The History of the Earth's Crust*. Princeton: Princeton University Press, pp. 73–89.
Vine, F. J., and Matthews, D. H., 1963. Magnetic anomalies over oceanic ridges. *Nature*, **199**, 947–949.
Vine, F. J., and Wilson, J. T., 1965. Magnetic anomalies over a young oceanic ridge off Vancouver Island. *Science*, **150**, 485–489.
Wegener, A., 1912. Die entstehung der kontinente. *Geologishe Rundschau*, **3**, 276–292.
Wilson, J. T., 1965. A new class of faults and their bearing on continental drift. *Nature*, **207**, 343–347.

Cross-references

Continental Drift
Continental Rifts
Lithosphere, Mechanical Properties
Lithosphere, Oceanic
Plate Driving Forces
Plates and Paleoreconstructions
Subduction Zones

SEDIMENTARY BASINS

Magdalena Scheck-Wenderoth
Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences
Sect. 4.4: Basin Analysis, Potsdam, Germany

Definition

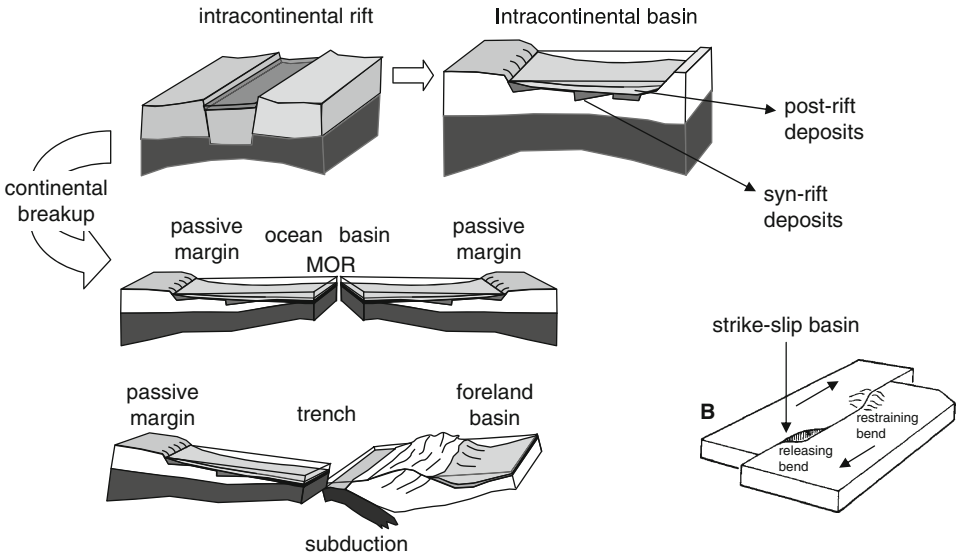
Sedimentary basins are regions of prolonged subsidence of the Earth's surface that provide the accommodation space for mineral and organic material (Allen and Allen, 2005). These deposits – the sedimentary rocks – are the record of the past geological history including tectonic events, climatic conditions, changes in sea level, and other environmental modifications. In addition, sedimentary basins are long-lived, low-temperature geo-reactors in which the accumulated material experiences a variety of transformations (Bjorlykke, 2010; Littke et al., 2008; Roure et al., 2009; Welte et al., 1997). As a result of these processes, basins contain our resources of fossil fuels, ground water, and inorganic commodities. Moreover, they are important reservoirs of heat and repositories for different socioeconomically relevant fluids such as CO_2.

Basin types

Basins can be classified in terms of their plate-tectonic setting. The plate-tectonic Wilson cycle (Figure 1) describes that the relative movements of plates on the sphere of the earth result in a closed loop of *continental rifting* and breakup, ocean basin development, closure of oceans at subduction zones and, in consequence of plate convergence, continental growth.

Basins thus can form as continental rifts that may evolve either into intracontinental basins or lead to continental breakup and ocean basin formation. New *oceanic lithosphere* (see entry *Lithosphere, Oceanic*) is produced along the Mid Ocean Ridges due to oceanic spreading. If oceanic lithosphere cools, it becomes denser and subsides in the oceanic basins that are bordered by passive continental margins. Of the latter "hot" or volcanic passive margins are distinguished from "cold" or nonvolcanic passive margins (White et al., 2003). If oceanic lithosphere cools beyond a specific threshold, it becomes too heavy to be sustained by the less dense asthenosphere and will finally descend back to the mantle at subduction zones. In these convergent settings, the descending plate is flexed downward and deep oceanic trenches develop above the down-going plate. Subduction may culminate in continent–continent collision if the oceanic lithosphere is entirely subducted and collisional fold and thrust belts form. Loading by collision-related fold and thrust belts also causes a downward flexure of the lithosphere and foreland basins to form. The Wilson Cycle may stop at any evolutionary step, because the causative forces cease to be effective. Accordingly, a continental rift not necessarily develops into an ocean basin, but may survive for hundreds of millions of years as an intracontinental basin (Heine et al., 2008; Littke et al., 2008). Finally, horizontal movements along strike-slip faults may also cause local extension and related pull-apart basins (Allen and Allen, 2005; Petrunin and Sobolev, 2008; Smit et al., 2010; Weber and group, 2009).

These different types of basins have a characteristic structure of the sediment fill as well as of the underlying crust and mantle lithosphere. This concerns the geometric configuration, the distribution of physical properties, and the resulting *isostatic* (see entry *Isostasy*) and thermal configuration. To assess the configuration of a basin, a wide spectrum of methods has to be integrated. Observations



Sedimentary Basins, Figure 1 Basins in their plate-tectonic setting.

obtained from field measurements, from *deep seismic imaging* (see entry *Deep Seismic Reflection and Refraction Profiling*), and from wells drilled into the basin fill, potential field data, and *heat flow* (see entry *Heat Flow Measurements, Continental*) data as well as new data sets from remote sensing need to be integrated with numerical models that simulate processes in basins at different scales.

Figure 2 shows exemplary crustal sections across the Norwegian passive margin and across the intracontinental Central European Basin System.

Basin-forming mechanisms

There are three main mechanisms leading to the formation of sedimentary basins (Allen and Allen, 2005)

1. Extensional stresses causing the lithosphere to thin
2. Flexure of the lithosphere in response to surface loading or compressive stresses
3. Viscous flow of the mantle causing positive or negative topography

All these mechanisms may act separately or in concert, and may affect the lithosphere over a range of spatial and temporal scales.

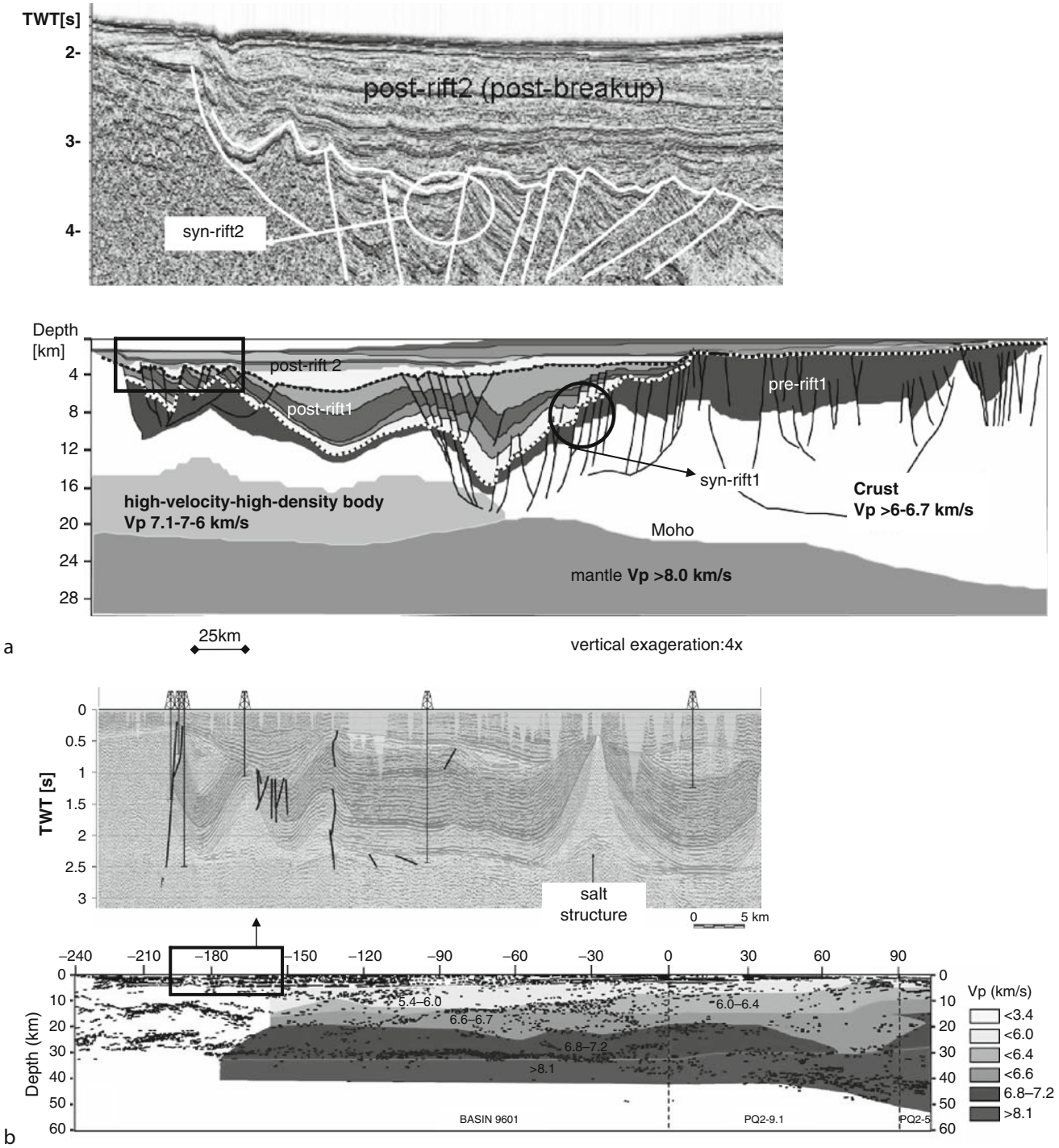
Conceptual models of basin formation

Conceptual models of basin formation (Figure 3) attempt to integrate geophysical and geological observations describing the geometry and the physical state of a basin into evolutionary schemes that explain basin evolution. These conceptual models also need to be consistent with data on the rheological behavior of rocks known from lab experiments as well as with data on the thermal and deformation history. All these conceptual models represent end-member scenarios and explain a large fraction of observed phenomena in specific basins.

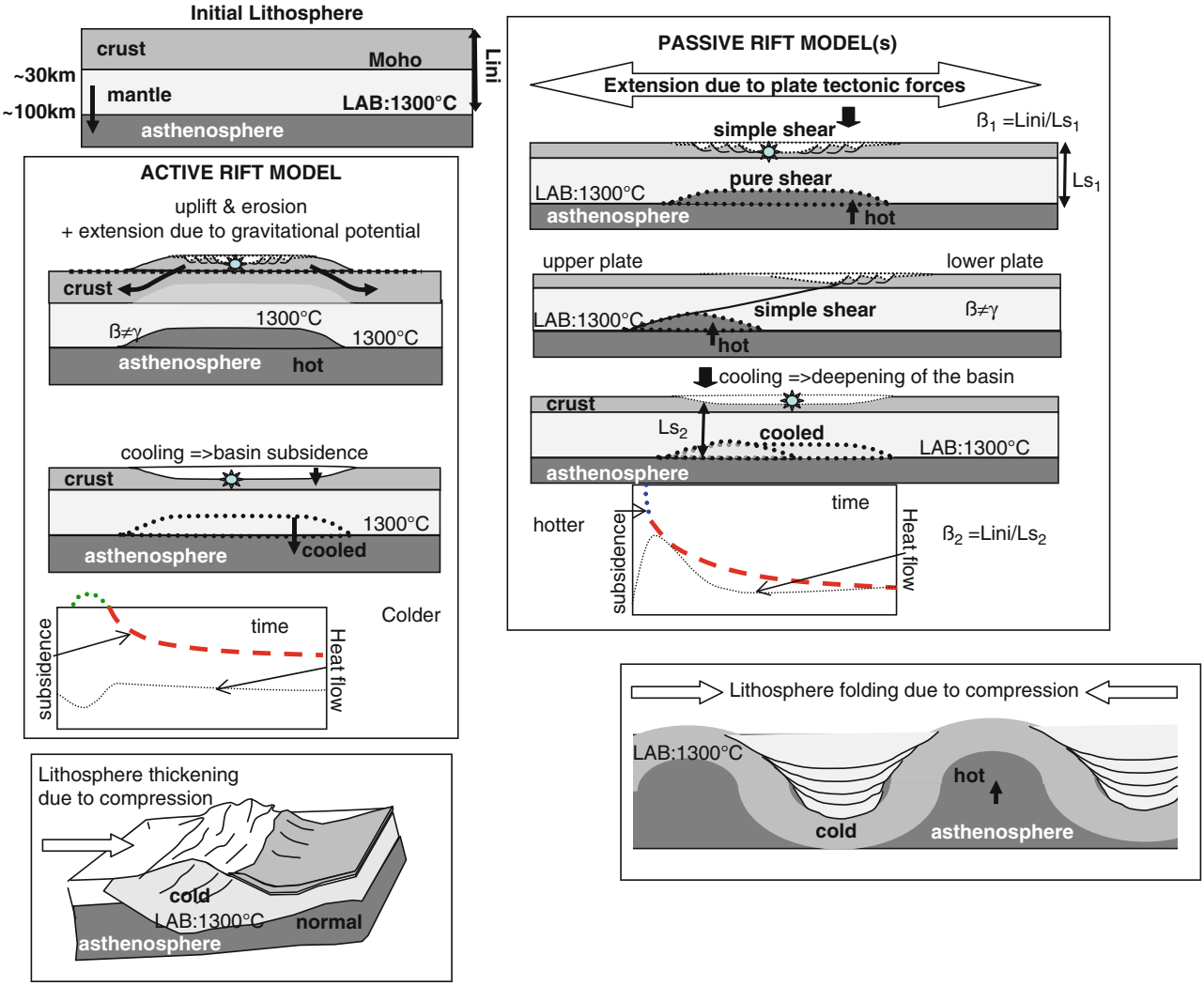
In the concept of *passive rifting* (Figure 3) the lithosphere is stretched in response to extensional tectonic stress. This may result in pure shear thinning (McKenzie, 1978), in simple shear thinning (Wernicke, 1981), or in a mixture of the two (Lister et al., 1986).

The *uniform stretching model* (McKenzie, 1978) predicts a crustal and lithosphere thickness proportional to the stretching factor ß where ß is the ratio between initial lithosphere thickness and the stretched lithosphere thickness. Lithosphere thinning results in a passive rise of the isotherm at the thermal lithosphere–asthenosphere boundary (Turcotte and Schubert, 2002). This, in turn, causes additional thermal subsidence after the stress ceases to be effective and cooling of the previously heated lithosphere causes a density increase. In this model, uniform and symmetric lithospheric thinning by ß takes place and is accommodated by brittle faulting in the upper crust as well as by ductile flow in the lower crust and upper mantle. The uniform stretching model predicts a 2-stage subsidence: (1) as a result of tectonic stretching – on a short timescale, (10–20 my), and (2) as a result of thermal subsidence – on a longer time scale (50–100 my). Structurally, the uniform stretching model predicts syn-rift extensional faults in the upper crust with syn-rift sediments overlain by onlapping post-rift sediments of the thermal subsidence phase. The thermal history predicted by the uniform stretching model includes an initial heat peak due to the rise of the isotherm at the LAB to a shallower position (and thus an increase of the thermal gradient within the lithosphere) followed by continuous cooling during thermal re-equilibration.

Depth-dependent stretching (Sclater and Christie, 1980) is a modification of the uniform stretching model



Sedimentary Basins, Figure 2 (**a**) *Lower panel*: Line drawing of a crustal scale integrated cross section with three generations of basins from the Norwegian passive margin modified after Faleide et al., 2008. *Upper panel*: Seismic close-up showing the breakup unconformity and pre-breakup extensional faulting during phase syn-rift 2. (**b**) *Lower panel*: Line drawing with superimposed seismic p-wave velocities from the intracontinental North German Basin modified after Bayer et al., 2002. *Upper panel*: Close-up with reflection seismic example illustrating polyphase subsidence with several sequences separated by unconformities and additional salt-tectonic deformation. (Modified after Mazur et al., 2005.)



Sedimentary Basins, Figure 3 Conceptual models of basin formation.

and accounts for different stretching magnitudes in the upper brittle part of the crust by a factor ß and the deeper lithosphere by a factor γ. In the *asymmetric stretching models* (Lister et al., 1986; Wernicke, 1981) crust- to lithosphere-scale listric faults accommodate large offsets in response to extensional stress and finally may lead to continental breakup with an "upper plate" margin developing from the hanging wall of the master detachment fault and a "lower plate" margin in the foot wall (Osmundsen and Ebbing, 2008; Sibuet et al., 2007; Zuber et al., 1986). In particular, nonvolcanic margins like the Galicia-Newfoundland (Sibuet et al., 2007) are commonly explained by this mechanism. Geodynamic modelling studies (Braun and van der Beek, 2004; Buck, 1991; Cloetingh et al., 1998; Huismans et al., 2005; Huismans and Beaumont, 2008; Kusznir and Ziegler, 1992; Lavier et al., 1999; Pascal and Cloetingh, 2002; Pérez-Gussinyé et al., 2006; Royden and Keen, 1980; Sclater and Christie, 1980; Simon et al., 2009; van Wijk, 2005; Wijns et al., 2005; Zuber et al., 1986) indicate that the nature of rheological coupling between the crust and the lithospheric mantle, local strain weakening and strengthening, and the rate of extension are the dominant factors controlling not only the subsidence evolution of basins but also if they develop in a symmetric or asymmetric style and in a narrow or wide rift mode.

In the concept of *active rifting*, basin initiation is caused by the impingement of a rising hot asthenospheric *mantle plume* (Allen and Allen, 2005; Burov and Cloetingh, 2009; Haxby et al., 1976; Turcotte and Schubert, 2002). The concept of active rifting is in fact a special case of basins initiated by *dynamic topography*. The emplacement of the *mantle plume* (see entry *Mantle Plumes*) causes uplift and induces extensional stresses at the flanks of the uplifted area due to gravitational potential. The active rise of the isotherm may lead to surface uplift above the

erosional level and the eroded material may be removed from the system. Density increase in the lithosphere due to subsequent cooling causes surface subsidence and the formation of a sedimentary basin in the previously uplifted area. As the process is transient, the system would re-equilibrate after the thermal anomaly and the related dynamic topography stop to exist. For *isostatic* reasons, net subsidence will only occur if either igneous underplating of the crust has taken place (light crust replaced by dense underplate) or if material is removed from the system by erosion during uplift. Systematic work assessing the contribution of the three processes to subsidence (extension due to gravity potential, crustal underplating, and erosion-related mass loss) is still missing. The active rift model predicts weak to missing syn-rift faulting, a syn-rift erosional unconformity, and a subsidence history corresponding predominantly to thermal subsidence. Actually, the active rift model predicts a subsidence history identical to a specific end-member case of (passive) depth-dependent stretching in which the mantle lithosphere is severely thinned whereas the crust is not. However, the active rift model predicts a colder heat flow scenario than the passive rift model, as the thermal gradient is not disturbed by an initial thinning of the lithosphere.

According to the concept of dynamic topography basins can also be initiated in response to downward flow of high density material in the convective mantle, a process causing *negative dynamic topography*. This mechanism could be related to detached subduction slabs sinking into the asthenospheric mantle and is discussed as an explanation for the subsidence of intracontinental basins that subside over several hundreds of millions of years at a very slow rate (Allen and Allen, 2005; Heine et al., 2008; Ritzmann and Faleide, 2009).

Two main conceptual models exist for *flexural basins* that form as response of the flexural rigidity of the lithosphere (Watts, 2001) to external loads: (1) downward flexure of the lithosphere due to vertical loading (Zhang and Bott, 2000) and (2) formation of crustal- to lithosphere-scale folds in response to horizontal compressive stress, a process also called *lithosphere buckling* (Burov and Cloetingh, 2009; Burov and Diament, 1992; Cloetingh and Burov, 2010). Both concepts predict a lack of syn-rift extensional faulting, a cold thermal scenario, and subsidence lasting only as long as the stress is effective (either vertical as load or horizontal as tectonic stress). Due to structural similarity, lithosphere buckling in response to far-field compressive tectonic stress (Burov and Cloetingh, 2009; Burov and Diament, 1992) can be mistaken as an expression of active rifting if the analyzed region is too small to capture the full wavelength of the buckling. Again, the different predicted subsidence and thermal histories can help to distinguish the two.

Vertical loading by sediments enhances subsidence in basins of all geodynamic settings (Watts, 2001). It has recently been proposed that mineral phase transitions may also play a role for subsidence history in different tectonic settings (Kaus et al., 2005).

The configuration of the sediment fill

The sediment fill can vary in thickness between a few to up to 20 km and is analyzed using stratigraphic methods (Allen and Allen, 2005; Catuneanu, 2006; Einsele et al., 1991; Vail et al., 1991) to assess the nature of sediments (sand, silt, shale, carbonates, evaporites), the depositional environment (continental or marine), and the climatic conditions prevailing during the time of deposition. Subsidence and deposition may alternate with nondeposition due to sea level fall or tectonic surface uplift, and lead to stratigraphic gaps separating different sequences (Figure 2). In addition, the rate of subsidence and sedimentation as well as of uplift and erosion may change with time. Stratigraphic gaps and changes in depositional conditions are expressed as unconformities separating sedimentary sequences the analysis of which is required to understand the history of a basin.

Progressive sedimentation and subsidence goes along with changes in physical properties of the deposited sediments. The two most prominent changes are (1) an increase in temperature and (2) a decrease in porosity with increasing burial depth. The initially high porosity of each layer decreases as the load of the covering deposits increases and fluids filling the initial pores are expelled. Also, chemical compaction may lead to a loss in porosity. With decreasing porosity, physical properties like density, seismic velocity, or thermal conductivity increase.

Stratigraphic analysis is performed in field studies if the fill of a basin is accessible in outcrops due to postdepositional uplift. Most of the sedimentary basins, however, extend by their very nature (areas of enduring subsidence) into the subsurface and to large depths. Accordingly, subsurface analytical methods are required to assess their configuration. The most important of these methods is *seismic imaging*. *Seismic reflections* (see entry *Deep Seismic Reflection and Refraction Profiling*) originate from the contrast in impedance at discontinuities in the subsurface that can represent stratigraphic boundaries, and unconformities or changes in attributes of a specific layer. The reflection seismic technique provides a picture of these interfaces in a two-way-travel time that is constrained by wells. Wells provide information on the absolute depth position of stratigraphic horizons and enable the conversion of seismic travel times to depth. Furthermore, wells yield data on the thermal gradient with depth as well as on changes in lithology and petrophysical properties. Properties such as composition, compaction, porosity, permeability, density, thermal and hydraulic conductivity, *radiogenic heat production* (see entry *Radiogenic Heat Production of Rocks*), and content and maturation of organic matter are measured using *Geophysical Well Logging* methods.

Apart from recording changes in relative sea level or climatic conditions, the configuration of the sediment fill may be influenced by deformation (Kley et al., 2008). Accordingly, sediments can be faulted, folded in response to tectonic forces, or deformed due to halokinetic movements (Hudec and Jackson, 2007; Scheck-Wenderoth

et al., 2008). This deformation can take place pre-, syn-, or postdepositionally, each of which is expressed by characteristic structural styles.

In particular, basin initiation in extensional plate-tectonic conditions is typically associated with normal faulting in the upper crust leading to a thinning of the lithosphere and to the accumulation of syn-rift deposits. After the tectonic stresses have declined, basins experience a phase of post-rift thermal subsidence and may even experience late basin inversion in response to compressive stress (Cooper and Williams, 1989; Mazur et al., 2005; Sandiford, 1999) (Figure 2).

Finally, there is a feedback between surface processes and deep mechanisms in the sense that *isostatic* loading by sediments will increase total subsidence and erosion of basin flanks will cause flexural rebound (Braun, 2005; Braun et al., 2008; Braun and van der Beek, 2004; Garcia-Castellanos, 2002; Kounov et al., 2007; Simoes et al., 2010; Van der Beek, 2007; Willenbring and von Blanckenburg, 2010).

The configuration of the crust beneath sedimentary basins

The different mechanisms responsible for the thinning of the lithosphere leave characteristic traces that can be detected with geophysical observations, which, however, are not always leading to unique interpretations. The most prominent attribute of the crust beneath sedimentary basins is that it is thinned compared to neighboring regions. Deep seismic data imaging of the preserved thickness of the crust helps to determine the amount of the stretching factor ß as the ratio between the initial crustal thickness as often preserved outside the basin margins and the crustal thickness beneath the basin. Crustal thinning can vary considerably across a basin with values of ß largest in the area of strongest thinning. The average initial thickness of the crust may reach up to 50 km in cratonic areas, around 30 km in Phanerozoic continental domains and less than 10 km below oceans.

As the crystalline crust is characterized by a higher density and higher velocities of seismic waves than the sediments, the interface between the two can be detected. Accordingly, the *interpretation of gravity anomalies* (see entry *Gravity Anomalies, Interpretation*) and the analysis of the observed variation of seismic velocities are key to evaluate the crustal structure. Refraction seismic techniques are used to detect variations in seismic p-wave velocities at depth whereas reflection seismic methods are useful to evaluate the changes in reflectivity of the deeper crust.

Seismic p-wave velocities (Vp) increase to values >6 km/s in the upper crystalline crust and can range up to 7.4 km/s (Christensen and Mooney, 1995) depending on the composition of the crust. There is a negative correlation between the Quartz content of crystalline rocks and seismic p-wave velocity as well as density. Silicic (Quartz-rich) rocks are characterized by smaller velocities and lower densities than mafic rocks (Table 1). Seismic shear wave velocities are also higher for crystalline than for sedimentary rocks, but are also dependent of additional parameters (Christensen and Mooney, 1995). Accordingly, the Vp/Vs-ratio is an additional parameter helping to interpret crustal structure (Afonso et al., 2010; Mjelde et al., 2003). As also the magnetic properties are different for mafic (Fe-rich) and silicic (Fe-poor) rocks, *the interpretation of magnetic anomalies* is an additional technique for the evaluation of the crustal structure beneath basins.

At the crust–mantle boundary (Moho), often a characteristic reflection and a sudden increase of seismic p-wave

Sedimentary Basins, Table 1 Overview on physical properties of rocks relevant to sedimentary basins: *seismic properties* (After Christensen and Mooney, 1995), *thermal properties* (After Artemieva, 2006; Cermak and Rybach, 1982; Fernandez et al., 2005; Förster and Förster, 2000; Ritter et al., 2004; Scheck-Wenderoth and Maystrenko, 2008)

	Dominant lithology	Thermal conductivity λ [W/mK]	Heat production S [mW/m^3]	Density [kg/m^3]	p-wave velocity [km/s]
	Sea water	0.563	0	1,030	1.48
S	Uncompacted fine-grained siliciclastics	1.2 (2.0)	1	1,950	2.05
E	Slightly compacted fine-grained siliciclastics	1.8 (2.2)	1	2,200	2.2
D	Medium compacted fine-grained siliciclastics	2.0 (2.2)	1	2,250	2.2
I	Compacted fine-grained siliciclastics	2.1 (2.3)	1	2,550	3.5
M	Strongly compacted fine-grained siliciclastics	2.1 (2.3)	1	2,630	4.3
E	Compacted siliciclastics with carbonates	2.1 (2.5)	0.9	2,640	5.5–6.0
N	Carbonates:chalk	1.7	0.6	2,000	3.4
T	Carbonates: limestones, dolomites	2.4	0.6	2,600	6.0
S	Rock salt	3.5 (5.4)	0.1	2,150	4.0–5.5
Crust	Granites and gneisses	2.7 (3.2)	0.8	2,790	6.0–6.7
Continent	Mafic granulites/gabbros/eclogites	2.6 (3.0)	0.3	3,120	6.8–7.0
	Gabbros	2.6 (3.0)	0.5	3,150	7.1–7.6
	Basalt	1.8 (2.1)	0.4	2,580	4.0–5.0
Crust	Sheeted dikes/gabbroic intrusions	2.3 (2.6)	0.2	2,890	5.0–6.7
Ocean	Gabbro	2.3 (3.0)	0.2	3,150	7.1–7.6
Mantle	Peridotite continent	3.95 (3.2)	0.03	3,330	8.0–8.3
	Peridotite ocean	3.95 (3.2)	0.03	3,180	7.8

velocity to values >8 km/s is observed in deep seismic data marking the transition to the lithospheric mantle (Bauer et al., 2000; DEKORPBASIN and Group, 1999; Fowler, 1996; Hirsch et al., 2009a; Levander et al., 2006; Meissner and Group, 1991; Thybo and Nielsen, 2009; Turcotte and Schubert, 2002; Weber and group, 2009). This is related to the change in composition from various crustal rocks to the mantle consisting mainly of peridotite.

As less dense sediments replace a denser crystalline crust, sedimentary basins should be characterized by a negative Bouguer anomaly with highest amplitude in the basin center after re-equilibration of the isotherm. Many basins, however, are characterized by a long-wavelength negative Bouger anomaly with positive anomalies superimposed. This observation, together with strong variations in the crustal velocity structure beneath sedimentary basins challenge the classical concepts of crustal thinning and related Moho uplift in that they commonly occur in concert with a flat Moho (Thybo and Nielsen, 2009).

Crustal bodies with seismic velocities >7 km, generally referred to as "high-velocity bodies," are observed in the lower crust beneath many intracontinental rift structures as, for example, the Baikal and the Kenya Rifts (Thybo and Nielsen, 2009), beneath intracontinental basins as, for example, in the Danish Basin (Nielsen and Thybo, 2006), and beneath extended passive margins as, for example, at the conjugate margins of the North and South Atlantic (Bauer et al., 2000; Contrucci et al., 2004; Faleide et al., 2008; Franke et al., 2007; Hirsch et al., 2009a; Mjelde et al., 2002; Mjelde et al., 2005; Sibuet et al., 2007). Gravity analysis indicates that these bodies are also characterized by higher densities than the average crust (Fernàndez et al., 2010, 2005; Franke et al., 2007; Hirsch et al., 2009a; Maystrenko and Scheck-Wenderoth, 2009; Osmundsen and Ebbing, 2008).

Differential thinning with depth (Huismans et al., 2005; Huismans and Beaumont, 2008; Kusznir and Ziegler, 1992; Lavier et al., 1999; Lavier and Steckler, 1997; Pascal and Cloetingh, 2002; Sclater and Christie, 1980; Steckler and Watts, 1978), emplacement of magmatic material in the crust during the thinning process or magmatic underplating (Thybo and Nielsen, 2009; Burg and Gerya, 2008) are some possible processes during basin formation and evolution that can be responsible for the actual configuration of the crust beneath a specific basin. The relative contributions of the processes vary strongly for different basins, and, accordingly, their relative importance in general terms is still under debate.

Moreover, structural inheritance may have a fundamental role in defining the crustal and lithosphere structure. Successive suturing of different plates in Earth history may result in a mosaic of crustal domains with contrasting physical properties, possibly also including lower crustal bodies predating the rifting process (Ebbing et al., 2009; Faleide et al., 2008; Plomerová and Babuska, 2010; Scheck-Wenderoth and Lamarche, 2005; van Wijk, 2005; Vauchez et al., 1998). In basins developing on such a substrate such older rheological discontinuities may be reactivated to localize deformation in areas of reduced strength, thus also facilitating discontinuous stretching with depth.

The configuration of the mantle lithosphere

Much less is known on the configuration of the mantle lithosphere below sedimentary basins due to the limited amount of direct observations. Though reflections and lateral variations in p-wave velocities in the lithospheric mantle have been observed in deep seismic data, the energy used in active seismic experiments is generally not sufficient to reach depth intervals below the Moho. Also, the gravity signal from the lithospheric mantle is difficult to isolate from the cumulative signal of the entire lithosphere. Recent development in the acquisition and evaluation of passive *seismological* data (Dalton and Faul, 2010; Fishwick, 2010; Geissler et al., 2010; Heintz and Kennett, 2005; Hieronymus and Goes, 2010; Plomerová and Babuska, 2010; Priestley et al., 2006; Ritzmann and Faleide, 2009; Zhang and Lay, 1996), electromagnetic and *magnetotelluric* methods (see entries *Magnetotelluric Interpretation*; *Magnetotelluric Data Processing*) (Jones et al., 2010), together with remote sensing gravity observation (Kaban et al., 2003; Schotman et al., 2009), as well as geochemical data (O'Reilly and Griffin, 2010; Trumbull et al., 2002; Wang, 2010) and thermal studies (Artemieva, 2006, 2009; Hasterok and Chapman, 2007; Hieronymus and Goes, 2010) indicates that the lithospheric mantle is less homogenous than previously thought. Apart from strong differences in thickness, also considerable lateral variations in surface *heat flow* (see entry *Heat Flow Measurements, Continental*) are observed, with a positive correlation between the two. The lithosphere-asthenosphere boundary (LAB) may be located deeper than 250 km in cratonic areas, lies at around 100 km beneath Phanerozoic continental domains and old oceans, and is close to the seafloor at Mid Ocean Ridges.

Agreement is established that the LAB is a fundamental boundary in plate-tectonic theory that separates the rigid plates from ductile convecting material below the plates (Artemieva, 2009; Eaton et al., 2009; Jones et al., 2010). Three broad definitions in terms of a mechanical boundary layer, a thermal boundary layer, and a chemical boundary layer are based on different types of data and partly in geometrical conflict. For the evolution of sedimentary basins, the depth of the thermal LAB is especially relevant as it determines the thermal and mechanical state of the lithosphere subjected to any of the basin-forming mechanisms. Systematic mapping of lithosphere thickness beneath sedimentary basins is, however, still lacking. The thermal LAB is interpreted as an isotherm of about 1,300°C and its depth corresponds to the depth where a continental conductive geotherm intersects the mantle adiabat. This is corroborated by heat flow inversion studies in *continental lithosphere* (see entry *Lithosphere, Continental*) (Artemieva, 2009), by combined thermal and gravity

modelling (Fernàndez et al., 2010, 2005; Hasterok and Chapman, 2007; Hyndman et al., 2009; Scheck-Wenderoth and Maystrenko, 2008) and by cooling models in *oceanic lithosphere* (see entry *Lithosphere, Oceanic*) (Crosby et al., 2006; McKenzie, 1978; Sclater, 2003; Stein and Stein, 1992).

Isostatically, thinning of the lithospheric mantle results in net surface uplift as heavier lithospheric mantle is replaced by less dense asthenospheric material (active rift model). Only if the interaction with surface processes takes place (erosion), net subsidence takes place due to subsequent cooling of the asthenospheric material.

Heat flow in sedimentary basins

The thermal field in sedimentary basins has been of primary interest in the exploration of fossil fuels (Welte et al., 1997) and gains increasing importance for the use of geothermal energy (Huenges, 2010). Moreover, together with composition, temperature is a main controlling factor for the rheological behavior of the lithosphere and accordingly its deformation. The hotter the lithosphere, the weaker is its rheology and the easier it is thinned to form a sedimentary basin. Accordingly, the main controlling factors for the thermal state of a basin are its plate-tectonic setting and its evolutionary stage.

In terms of observables, surface *heat flow* measurements and temperatures measured in wells characterize the present-day thermal state of a basin, whereas the maturity of organic matter (Welte et al., 1997) and thermochronological data (Andriessen, 1995; Braun, 2005; Braun and van der Beek, 2004; Kounov et al., 2007; Simoes et al., 2010; Van der Beek, 2007; Willenbring and von Blanckenburg, 2010) provide a record of the thermal history.

Different families of integrated process-oriented models attempt to reproduce these observables and indicate that several processes contribute to the heat flow in sedimentary basins. There is a first-order contrast in thermal properties between sedimentary and crystalline rocks, in that sediments are thermally less conductive and produce less radiogenic heat than crystalline crustal rocks (Table 1). The amount of heat entering the basin at the base of the sediment fill depends on the thickness and composition of the crystalline crust as well as on the depth of the thermal LAB. The mafic components of the crust and the lithospheric mantle are characterized by a high thermal conductivity but low *radiogenic heat production*. Accordingly, the shallower the LAB and the thicker and radiogenic the crystalline crust, the more heat arrives at the base of the sediments. Due to their higher porosities, the thermally low-conductive sediments act as a thermal blanket, causing heat storage in the basin (Cacace et al., 2010; Theissen and Rüpke, 2009; Van Wees et al., 2009). In addition, the sediments contribute a modest but, in the sum, considerable amount of radiogenic heat to the system. Also within the sediment fill, the thermal properties may vary (Table 1) with the thermal conductivity of salt being two times larger than that of clastic sediments. The upper part of sedimentary basins may store paleoclimatic signals of previous glaciations as present-day permafrost down to more than 700 m depth attest.

In response to the distribution of thermal parameters long-wavelength variations in temperatures in sedimentary basins (scale of hundreds of kilometers) are determined by the crustal structure and composition as well as by the thickness of the lithosphere (Cacace et al., 2010; Hasterok and Chapman, 2007; Hyndman et al., 2009; Scheck-Wenderoth and Maystrenko, 2008; Sclater, 2003). In contrast, the short-wavelength pattern of temperature distribution (scale of kilometers) is controlled by the internal configuration of the sediment fill.

Heat transfer by conduction is assumed to dominate the transport of heat in the lithosphere in contrast to the *convecting mantle* (see entry *Mantle Convection*). While geological and geochemical data proved that additional hydrothermal heat transport takes place along faults in the lithosphere, there is an ongoing debate on how far the thermal field in sedimentary basins is additionally influenced by convective heat transport due to moving pore fluids. The inherent vertical anisotropy due to the layered nature of sediments counteracts the development of free convection on a basin-wide scale (Bjorlykke, 2010) but local convection is indicated by geochemical data from hydrogeological studies and models of coupled fluid and heat transport (Magri et al., 2009).

Summarizing, the heat flow regime in a specific basin may vary spatially though some ranges can be given for different types of basins (Allen and Allen, 2005): Typically, the surface *heat flow* in intracontinental basins varies between 40 and 70 mW/m^2, can be up to 150 mW/m^2 in active rifts or close to volcanic arcs, and can reach values higher than 180 mW/m^2 at mid ocean ridges and oceanic rifts. In oceanic basins, surface *heat flow* decreases with increasing distance from the mid ocean ridges according to the cooling of the lithosphere with age (Parsons and Sclater, 1977; Stein and Stein, 1992). The surface *heat flow* at passive margins depends on the age of the adjacent oceanic lithosphere. At young margins as in the 55-my-old North Atlantic, the surface heat flow is still controlled by the cooling of the oceanic lithosphere that is considerably thinner than the lithosphere of the continental margin. This step in the thermal LAB is consistent with observed heat flow values increasing from 45 mW/m^2 at the continental side to 80 mW/m^2 at the oceanic side of the margin (Ritter et al., 2004; Scheck-Wenderoth and Maystrenko, 2008). At older passive margins as in the 130-my-old South Atlantic, an opposite trend is observed (Goutorbe and Bonneville, 2008; Hirsch et al., 2009b) with about 45 mW/m^2 at the oceanic side of the margin to up 65 mW/m^2 at the continental side. As the oceanic lithosphere had sufficient time to cool and thicken, the depth of the thermal LAB is continuous and similar between the continental and the oceanic part of the system. Instead, the radiogenic heat contribution of the *continental crust* (see entry *Earth's*

Structure, Continental Crust) thickening toward the continent becomes the controlling factor for heat input from the lithosphere.

Flexural basins are generally colder than extensional basins with average *heat flow* values around 40 mW/m^2 in oceanic trenches and foreland basins. The thermal signature of strike-slip basins is highly variable as it depends on the attributes of the lithosphere on which they develop and if hydrothermal heat transport takes place along the fault zones.

Summary

The *geodynamics* of sedimentary basins results from the interaction of a variety of processes acting on different spatial and temporal scales. Geophysical and geological observations indicate that intra-plate stress and heat input from the asthenospheric mantle are first-order controlling factors. On the *continental lithosphere* (see entry *Lithosphere, Continental*), structural inheritance and its impact on rheology is a third major player controlling the development of sedimentary basins. Intra-plate stress determines if an extensional or a compressional basin is formed. Temperature dominantly controls the rheological behavior of the lithosphere in addition to composition. These two parameters determine if the lithosphere is thinned uniformly, depth dependently, and in a symmetric or asymmetric mode. The more layered the lithosphere rheology is, the more discontinuous and asymmetric the stretching process will be. In contrast, a lithosphere deforms in a uniform stretching mode if the crust is strongly coupled to the mantle and no significant vertical or horizontal rheology contrasts are present. In both cases, the magnitude and rates of effective stress as well as the magnitude of the heat anomaly determine if stretching takes place slow enough to allow for cooling related strain hardening or fast enough to result in continental breakup.

These deep factors interact continuously with surface processes such as deposition or erosion of sediments influenced by climatic conditions. Deposition leads to *isostatic* loading and enhanced subsidence whereas erosion results in isostatic unloading and enhanced uplift.

Bibliography

Afonso, J. C., et al., 2010. On the Vp/Vs-Mg# correlation in mantle peridotites: implications for the identification of thermal and compositional anomalies in the upper mantle. *Earth and Planetary Science Letters*, **289**(3–4), 606–618.

Allen, P., and Allen, J., 2005. *Basin Analysis. Principles and Applications*. Oxford: Blackwell Science, p. 560.

Andriessen, P. A. M., 1995. Fission-track analysis: principles, methodology and implications for tectono-thermal histories of sedimentary basins, orogenic belts, and continental margins. *Geologie en Mijnbouw*, **74**, 1–12.

Artemieva, I. M., 2006. Global 1° × 1° thermal model TC1 for the continental lithosphere: implications for lithosphere secular evolution. *Tectonophysics*, **416**(1–4), 245–277.

Artemieva, I. M., 2009. The continental lithosphere: reconciling thermal, seismic, and petrologic data. *Lithos*, **109**, 23–46.

Bauer, K., et al., 2000. Deep structure of the Namibia continental margin as derived from integrated geophysical studies. *Journal of Geophysical Research*, **105**(B11), 25,829–25,853.

Bayer, U., Grad, M., Pharaoh, T. C., Thybo, H., Guterch, A., Banka, D., Lamarche, J., Lassen, A., Lewerenz, B., Scheck, M., and Marotta, A. M., 2002. The southern margin of the East European Craton: new results from seismic sounding and potential fields between the North Sea and Poland. *Tectonophysics*, **360**, 301–314.

Bjorlykke, K., 2010. *Heat Transport in Sedimentary Basins, Petroleum Geoscience*. Berlin: Springer, pp. 253–257.

Braun, J., 2005. Quantitative constraints on the rate of landform evolution derived from low-temperature thermochrohonology, low-temperature thermochronology: techniques, interpretations, and applications. *Reviews in Mineralogy and Geochemistry*, **58**, 351–374.

Braun, J., and van der Beek, P., 2004. Evolution of passive margin escarpments: what can we learn from low-temperature thermochronology? *Journal of Geophysical Research-Earth Surface*, **109**, F04009.

Braun, J., et al., 2008. DOUAR: a new three-dimensional creeping flow numerical model for the solution of geological problems. *Physics of the Earth and Planetary Interiors*, **171**(1–4), 76–91.

Buck, W. R., 1991. Modes of continental lithospheric extension. *Journal of Geophysical Research*, **96**(B12), 20.161–20.178.

Burg, J. P., and Gerya, T. V., 2008. Modelling intrusion of mafic and ultramafic magma into the continental crust: numerical methodology and results. *Bollettino Della Societa Geologica Italiana*, **127**(2), 199–203.

Burov, E., and Cloetingh, S., 2009. Controls of mantle plumes and lithospheric folding on modes of intraplate continental tectonics: differences and similarities. *Geophysical Journal International*, **178**(3), 1691–1722.

Burov, E. B., and Diament, M., 1992. Flexure of the continental lithosphere with multilayered rheology. *Geophysical Journal International*, **109**, 449–468.

Cacace, M., Kaiser, B. O., Lewerenz, B., and Scheck-Wenderoth, M., 2010. Geothermal energy in sedimentary basins: what we can learn from regional numerical models. *Chemie Der Erde-Geochemistry*, **70**, 33–46.

Catuneanu, O., 2006. *Principles of Sequence Stratigraphy*. New York: Elsevier, p. 386.

Cermak, V., and Rybach, L., 1982. Thermal properties: Thermal conductivity and specific heat of minerals and rocks. In Angenheister, G. (ed.), *Landolt-Börnstein, New Series, Geophysics and Space Research*. Berlin: Springer, pp. 305–343.

Christensen, N. I., and Mooney, W. D., 1995. Seismic velocity structure and composition of the continental crust: a global view. *Journal of Geophysical Research*, **100**(B7), 9761–9788.

Cloetingh, S., and Burov, E., 2010. Lithospheric folding and sedimentary basin evolution: a review and analysis of formation mechanisms. *Basin Research*, doi:10.1111/j.1365-2117.2010.00490.x.

Cloetingh, S., Boldreel, L. O., Larsen, B. T., Heinesen, M., and Mortensen, L., 1998. Tectonics of sedimentary basin formation: models and constraints. *Tectonophysics*, **300**(I–IV), 1–11.

Contrucci, I., et al., 2004. Deep structure of the West African continental margin (Congo, Zaire, Angola), between 5 degrees S and 8 degrees S, from reflection/refraction seismics and gravity data. *Geophysical Journal International*, **158**(2), 529–553.

Cooper, M. A., and Williams, G. D., 1989. *Inversion Tectonics. Geological Society Special Publication Classics*. London: The Geological Society of London.

Crosby, A. G., McKenzie, D., and Sclater, J. G., 2006. The relationship between depth, age and gravity in the oceans. *Geophysical Journal International*, **166**(2), 553–573.

Dalton, C. A., and Faul, U. H., 2010. The oceanic and cratonic upper mantle: clues from joint interpretation of global velocity and attenuation models. *Lithos*, **120**(1–2), 160–172.
DEKORP-BASIN and Group, R., 1999. Deep crustal structure of the Northeast German basin: new DEKORP-BASIN'96 deep-profiling results. *Geology*, **27**(1), 55–58.
Eaton, D. W., et al., 2009. The elusive lithosphere-asthenosphere boundary (LAB) beneath cratons. *Lithos*, **109**(1–2), 1–22.
Ebbing, J., Gernigon, L., Pascal, C., Olesen, O., and Osmundsen, P. T., 2009. A discussion of structural and thermal control of magnetic anomalies on the mid-Norwegian margin. *Geophysical Prospecting*, **57**(4), 665–681.
Einsele, G., Ricken, W., and Seilacher, A., 1991. Cycles and Events in Stratigraphy - Basic Concepts and Terms. In Einsele, G., Ricken, W., and Seilacher, A. (eds.), *Cycles and Events in Stratigraphy*. Berlin: Springer, pp. 1–19.
Faleide, J. I., et al., 2008. Structure and evolution of the continental margin off Norway and the Barents Sea. *Episodes*, **31**(1), 82–91.
Fernandez, M., et al., 2005. Lithospheric structure of the mid-Norwegian margin; comparison between the More and Voring margins. *Journal of the Geological Society of London*, **162**(6), 1005–1012.
Fernàndez, M., Afonso, J. C., and Ranalli, G., 2010. The deep lithospheric structure of the Namibian volcanic margin. *Tectonophysics*, **481**(1–4), 68–81.
Fishwick, S., 2010. Surface wave tomography: imaging of the lithosphere-asthenosphere boundary beneath central and southern Africa? *Lithos*, **120**(1–2), 63–73.
Förster, A., and Förster, H.-J., 2000. Crustal composition and mantle heat flow: implications from surface heat flow and radiogenic heat production in the Variscian Erzbgebirge. *Journal of Geophysical Research*, **105**(B12), 917–938.
Fowler, C. M. R., 1996. *The Solid Earth*. Cambridge: Cambridge University Press, p. 472.
Franke, D., Neben, S., Ladage, S., Schreckenberger, B., and Hinz, K., 2007. Margin segmentation and volcano-tectonic architecture along the volcanic margin off Argentina/Uruguay, South Atlantic. *Marine Geology*, **244**(1–4), 46–67.
Garcia-Castellanos, D., 2002. Interplay between lithospheric flexure and river transport in foreland basins. *Basin Research*, **14**(2), 89–104.
Geissler, W. H., Sodoudi, F., and Kind, R., 2010. Thickness of the central and eastern European lithosphere as seen by S receiver functions. *Geophysical Journal International*, **181**(2), 604–634.
Goutorbe, B. L. F., and Bonneville, A., 2008. The thermal regime of South African continental margins. *Earth and Planetary Science Letters*, **267**(1–2), 256–265.
Hasterok, D., and Chapman, D. S., 2007. Continental thermal isostasy: 1. Methods and sensitivity. *Journal of Geophysical Research*, **112**(B6), B06414.
Haxby, W. F., Turcotte, D. L., and Bird, J. M., 1976. Thermal and mechanical evolution of the Michigan Basin. *Tectonophysics*, **36**(1–3), 57–75.
Heine, C., Dietmar Müller, R., Steinberger, B., and Torsvik, T. H., 2008. Subsidence in intracontinental basins due to dynamic topography. *Physics of the Earth and Planetary Interiors*, **171**(1–4), 252–264.
Heintz, M., and Kennett, B. L. N., 2005. Continental scale shear wave splitting analysis: investigation of seismic anisotropy underneath the Australian continent. *Earth and Planetary Science Letters*, **236**(1–2), 106–119.
Hieronymus, C. F., and Goes, S., 2010. Complex cratonic seismic structure from thermal models of the lithosphere: effects of variations in deep radiogenic heating. *Geophysical Journal International*, **180**(3), 999–1012.
Hirsch, K. K., Bauer, K., and Scheck-Wenderoth, M., 2009a. Deep structure of the western South African passive margin – results of a combined approach of seismic, gravity and isostatic investigations. *Tectonophysics*, **470**(1–2), 57–70.
Hirsch, K. K., Scheck-Wenderoth, M., Maystrenko, Y., Sippel, J., and Anonymous, 2009. A lithospheric 3D temperature study from the South Atlantic. *Geophysical Research Abstracts*, **11**, EGU2009-8053.
Hudec, M. R., and Jackson, M. P. A., 2007. Terra infirma: understanding salt tectonics. *Earth Science Reviews*, **82**(1–2), 1–28.
Huenges, E., 2010. *Geothermal Energy Systems: Exploration, Development and Utilization*. Weinheim: Wiley-VCH, p. 464.
Huismans, R. S., and Beaumont, C., 2008. Complex rifted continental margins explained by dynamical models of depth-dependent lithospheric extension. *Geology*, **36**(2), 163–166.
Huismans, R. S., Buiter, S. J. H., and Beaumont, C., 2005. Effect of plastic-viscous layering and strain softening on mode selection during lithospheric extension. *Journal of Geophysical Research*, **110**, B02406.
Hyndman, R. D., Currie, C. A., Mazzotti, S., and Frederiksen, A., 2009. Temperature control of continental lithosphere elastic thickness, Te vs Vs. *Earth and Planetary Science Letters*, **277**(3–4), 539–548.
Jones, A. G., Plomerova, J., Korja, T., Sodoudi, F., and Spakman, W., 2010. Europe from the bottom up: a statistical examination of the central and northern European lithosphere-asthenosphere boundary from comparing seismological and electromagnetic observations. *Lithos*, **120**(1–2), 14–29.
Kaban, M. K., Schwintzer, P., Artemieva, I. M., and Mooney, W. D., 2003. Density of the continental roots; compositional and thermal contributions. *Earth and Planetary Science Letters*, **209**(1–2), 53–69.
Kaus, B. J. P., Connolly, J. A. D., Podladchikov, Y. Y., and Schmalholz, S. M., 2005. Effect of mineral phase transitions on sedimentary basin subsidence and uplift. *Earth and Planetary Science Letters*, **233**(1–2), 213–228.
Kley, J., Franzke, H.-J., Jähne, F., Krawczyk, C., Lohr, T., Reicherter, K., Scheck-Wenderoth, M., Sippel, J., Tanner, D., and van Gent, H., 2008. Strain and stress. In Littke, R. B., Baeyer, U., Gajewski, D., and Nelskamp, S. (eds.), *Dynamics of Complex Intracontinental Basins. The Central European Basin System*. Berlin: Springer, pp. 97–124.
Kounov, A., et al., 2007. Present denudation rates at selected sections of the South African escarpment and the elevated continental interior based on cosmogenic 3He and 21Ne. *South African Journal of Geology*, **110**(2/3), 235–248.
Kusznir, N. J., and Ziegler, P. A., 1992. The mechanics of continental extension and sedimentary basin formation: a simple-shear/pure-shear flexual cantilever model. *Tectonophysics*, **215**, 117–131.
Lavier, L. L., and Steckler, M. S., 1997. The effect of sedimentary cover on the flexural strength of continental lithosphere. *Nature*, **389**, 476–479.
Lavier, L. L., Buck, W. R., and Poliakov, A. N. B., 1999. Self-consistent rolling-hinge model for the evolution of large-offset low-angle normal faults. *Geology*, **27**(12), 1127–1130.
Levander, A., Niu, F., Lee, C.-T. A., and Cheng, X., 2006. Imag(in)ing the continental lithosphere. *Tectonophysics*, **416**(1–4), 167–185.
Lister, G. S., Etheridge, M. A., and Symonds, P. A., 1986. Datachement faulting and the evolution of passive margins. *Geology*, **14**, 246–250.
Littke, R., Bayer, U., Gajewski, D., and Nelskamp, S., 2008. *Dynamics of Complex Intracontinental Basins: The Central European Basin System*. Berlin: Springer.
Magri, F., et al., 2009. Salty groundwater flow in the shallow and deep aquifer systems of the Schleswig-Holstein area (North German Basin). *Tectonophysics*, **470**(1–2), 183–194.
Maystrenko, Y., and Scheck-Wenderoth, M., 2009. Density contrasts in the upper mantle and lower crust across the continent-ocean transition: constraints from 3-D gravity modelling at the

Norwegian margin. *Geophysical Journal International*, **179**(1), 536–548.
Mazur, S., Scheck-Wenderoth, M., and Krzywiec, P., 2005. Different modes of the late cretaceous-early tertiary inversion in the North German and Polish basins. *International Journal of Earth Sciences*, **94**(5–6), 782–798.
McKenzie, D., 1978. Some remarks on the development of sedimentary basins. *Earth and Planetary Science Letters*, **40**, 25–32.
Meissner, R., and Group, D. R., 1991. The DEKORP Surveys: Major results in tectonic and reflective styles. In Meissner, R., et al. (eds.), *Continental Lithosphere: Deep Seismic Reflections*. Washington, DC: American Geophysical Union, pp. 69–76.
Mjelde, R., et al., 2002. Lower crustal seismic velocity-anomalies; magmatic underplating or serpentinized peridotite? Evidence from the Voring Margin, NE Atlantic. *Marine Geophysical Researches*, **23**(2), 169–183.
Mjelde, R., et al., 2003. V (sub p)/V (sub s) ratio along the Voring Margin, NE Atlantic, derived from OBS data; implications on lithology and stress field. *Tectonophysics*, **369**(3–4), 175–197.
Mjelde, R., et al., 2005. Continent-ocean transition on the Vøring Plateau, NE Atlantic, derived from densely sampled ocean bottom seismometer data. *Journal of Geophysical Research*, **110**, B05101, doi:10.1029/2004JB003026.
Nielsen, L., and Thybo, H., 2006. Identification of crustal and upper mantle heterogeneity by modelling of controlled-source seismic data. *Tectonophysics*, **416**(1–4), 209–228.
O'Reilly, S. Y., and Griffin, W. L., 2010. The continental lithosphere-asthenosphere boundary: can we sample it? *Lithos*, **120**(1–2), 1–13.
Osmundsen, P. T., and Ebbing, J., 2008. Styles of extension offshore mid-Norway and implications for mechanisms of crustal thinning at passive margins. Tectonics, **27**(6), TC6016.
Parsons, B., and Sclater, J. G., 1977. An analysis of the variation of ocean floor bathymetry and heat flow with age. *Journal of Geophysical Research*, **82**(5), 803–827.
Pascal, C., and Cloetingh, S. A. P. L., 2002. Rifting in heterogeneous lithosphere: inferences from numerical modelling of the northern North Sea and the Oslo Graben. *Tectonics*, **21**(6), 1060–1071.
Pérez-Gussinyé, M., Morgan, J. P., Reston, T. J., and Ranero, C. R., 2006. The rift to drift transition at non-volcanic margins: insights from numerical modelling. *Earth and Planetary Science Letters*, **244**(1–2), 458–473.
Petrunin, A. G., and Sobolev, S. V., 2008. Three-dimensional numerical models of the evolution of pull-apart basins. *Physics of the Earth and Planetary Interiors*, **171**(1–4), 387–399.
Plomerová, J., and Babuska, V., 2010. Long memory of mantle lithosphere fabric – European LAB constrained from seismic anisotropy. *Lithos*, **120**(1–2), 131–143.
Priestley, K., McKenzie, D., and Debayle, E., 2006. The state of the upper mantle beneath southern Africa. *Tectonophysics*, **416** (1–4), 101–112.
Ritter, U., Zielinski, G. W., Weiss, H. M., Zielinski, R. L. B., and Sættem, J., 2004. Heat flow in the Vøring basin, mid-Norwegian shelf. *Petroleum Geoscience*, **10**, 353–365.
Ritzmann, O., and Faleide, J. I., 2009. The crust and mantle lithosphere in the Barents Sea/Kara Sea region. *Tectonophysics*, **470** (1–2), 89–104.
Roure, F., Cloetingh, S., Scheck-Wenderoth, M., and Ziegler, P.A., 2009. Achievements and challenges in sedimentary basin dynamics: a review. In Cloetingh, S., and Negendank, J., (eds.), *New Frontiers in Integrated Solid Earth Sciences.* International Year of Planet Earth. Springer, Dordrecht, pp. 145–233.
Royden, L., and Keen, C. E., 1980. Rifting process and thermal evolution of the continental margin of eastern Canada determined from subsidence curves. *Earth and Planetary Science Letters*, **51**, 343–361.
Sandiford, M., 1999. Mechanics of basin inversion. *Tectonophysics*, **305**, 109–120.
Scheck-Wenderoth, M., and Lamarche, J., 2005. Crustal memory and basin evolution in the Central European Basin System – new insights from a 3D structural model. *Tectonophysics*, **397** (1–2), 143–165.
Scheck-Wenderoth, M., and Maystrenko, Y., 2008. How warm are passive continental margins? A 3-D lithosphere-scale study from the Norwegian margin. *Geology*, **36**(5), 419–422.
Scheck-Wenderoth, M., Maystrenko, Y., Huebscher, C., Hansen, M., and Mazur, S., 2008. *Dynamics of Salt Basins. Dynamics of Complex Intracontinental Basins; the Central European Basin System*. Berlin: Springer, pp. 307–322.
Schotman, H. H. A., Vermeersen, L. L. A., Wu, P., Drury, M. R., and De Bresser, J. H. P., 2009. Constraints on shallow low-viscosity zones in Northern Europe from future GOCE gravity data. *Geophysical Journal International*, **178**(1), 65–84.
Sclater, J. G., 2003. *Heat Flow Under the Oceans, Plate Tectonics; an Insider's History of Modern Theory of the Earth*. Boulder: Westview Press.
Sclater, J. G., and Christie, P. A. F., 1980. Continental stretching: an explanation of the post-mid-cretaceous subsidence of the central North Sea basin. *Journal of Geophysical Research*, **85**(B7), 3711–3739.
Sibuet, J.-C., Srivastava, S., and Manatschal, G., 2007. Exhumed mantle-forming transitional crust in the Newfoundland-Iberia rift and associated magnetic anomalies. *Journal of Geophysical Research*, **112**(B6), B06105.
Simoes, M., Braun, J., and Bonnet, S., 2010. Continental-scale erosion and transport laws: a new approach to quantitatively investigate macroscale landscapes and associated sediment fluxes over the geological past. *Geochemistry Geophysics Geosystems*, **11**, Q09001.
Simon, K., Huismans, R. S., and Beaumont, C., 2009. Dynamical modelling of lithospheric extension and small-scale convection: implications for magmatism during the formation of volcanic rifted margins. *Geophysical Journal International*, **176** (Suppl.1), 327–350.
Smit, J., Brun, J. P., Cloetingh, S., and Ben-Avraham, Z., 2010. The rift-like structure and asymmetry of the Dead Sea Fault. *Earth and Planetary Science Letters*, **290**(1–2), 74–82.
Steckler, M. S., and Watts, A. B., 1978. Subsidence history and tectonic evolution of atlantic-type continental margins. *Earth and Planetary Science Letters*, **41**, 1–13.
Stein, C. A., and Stein, S., 1992. A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature (London)*, **359**(6391), 123–129.
Theissen, S., and Rüpke, L. H., 2009. Feedbacks of sedimentation on crustal heat flow: New insights from the Vøring Basin, Norwegian Sea. Basin Research, **22**(6), 976–990.
Thybo, H., and Nielsen, C. A., 2009. Magma-compensated crustal thinning in continental rift zones. *Nature*, **457**(7231), 873–876.
Torsvik, T. H., Rousse, S., Labails, C., and Smethurst, M. A., 2009. A new scheme for the opening of the South Atlantic ocean and the dissection of an Aptian salt basin. *Geophysical Journal International*, **177**(3), 1315–1333.
Trumbull, R. B., Sobolev, S. V., and Bauer, K., 2002. Petrophysical modeling of high seismic velocity crust at the Namibian volcanic margin. *Special Paper – Geological Society of America*, **362**, 221–230.
Turcotte, D. L., and Schubert, G., 2002. *Geodynamics*, 2nd edn. Cambridge: Cambridge University Press, p. 456.
Vail, P. R., Audemard, F., Bowman, S. A., Eisner, P. N., and Perez-Cruz, C., 1991. The Stratigraphic Signatures of Tectonics, Eustasy and Sedimentology - an Overview. In Einsele, G., Ricken, W., and Seilacher, A. (eds.), *Cycles and Events in Stratigraphy*. Berlin/Heidelberg: Springer, pp. 617–659.

Van der Beek, P., 2007. Thermochronological age – elevation profiles, denudation rates and relief development. *Geochimica Et Cosmochimica Acta*, **71**(15), A1055–A1055.
Van Wees, J. D., et al., 2009. Probabilistic tectonic heat flow modeling for basin maturation: assessment method and applications. *Marine and Petroleum Geology*, **26**(4), 536–551.
van Wijk, J. W., 2005. Role of weak zone orientation in continental lithosphere extension. *Geophysical Research Letters*, **32**, L02303.
Vauchez, A., Tommasi, A., and Barruol, G., 1998. Rheological heterogeneity, mechanical anisotropy and deformation of the continental lithosphere. *Tectonophysics*, **296**, 61–86.
Wang, Q., 2010. A review of water contents and ductile deformation mechanisms of olivine: implications for the lithosphere-asthenosphere boundary of continents. *Lithos*, **120**(1–2), 30–41.
Watts, A. B., 2001. *Isostasy and Flexure of the Lithosphere*. Cambridge: Cambridge University Press, p. 458.
Weber, M., and group, D.W., 2009. Anatomy of the Dead Sea Transform from lithospheric to microscopic scale. Reviews in Geophysics, **47**, RG2002, doi:10.1029/2008RG000264.
Welte, D. H., Horsfield, B., and Baker, D. R. (eds.), 1997. *Petroleum and Basin Evolution*. Berlin: Springer, p. 535.
Wernicke, B., 1981. Low-angle normal faults in the Basin and Range Province: nappe tectonics in an extending orogen. *Nature*, **291**, 645–648.
White, N., Thompson, M., and Barwise, T., 2003. Understanding the thermal evolution of deep-water continental margins. *Nature*, **426**(6964), 334–343.
Wijns, C., Weinberg, R., Gessner, K., and Moresi, L., 2005. Mode of crustal extension determined by rheological layering. *Earth and Planetary Science Letters*, **236**(1–2), 120–134.
Willenbring, J., and von Blanckenburg, F., 2010. Long-term stability of global erosion rates and weathering during late-Cenozoic cooling. *Nature*, **465**, 211–214.
Zhang, G.-B., and Bott, M. H. P., 2000. Modelling the evolution of asymmetrical basins bounded by high-angle reverse faults with application to foreland basins. *Tectonophysics*, **322**, 203–218.
Zhang, Y. S., and Lay, T., 1996. Global surface wave phase velocity variations. *Journal of Geophysical Research, Solid Earth and Planets*, **101**(B4), 8415–8436.
Zuber, M. T., Parmentier, E. M., and Fletcher, R. C., 1986. Extension of continental lithosphere: a model for two scales of basin and range deformation. *Journal of Geophysical Research*, **91**(B5), 4826–4838.

Cross-references

Continental Rifts
Deep Seismic Reflection and Refraction Profiling
Earth's Structure, Continental Crust
Geodynamics
Geophysical Well Logging
Gravity Anomalies, Interpretation
Heat Flow, Continental
Heat Flow, Seafloor: Methods and Observations
Isostasy
Lithosphere, Continental
Lithosphere, Mechanical Properties
Lithosphere, Oceanic
Lithosphere, Oceanic: Thermal Structure
Magnetic Anomalies, Interpretation
Magnetotelluric Interpretation
Mantle Convection
Mantle Plumes
Radiogenic Heat Production of Rocks
Seafloor Spreading
Seismic Imaging, Overview
Seismic Properties of Rocks
Seismology, Global Earthquake Model
Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat
Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity

SEISMIC ANISOTROPY

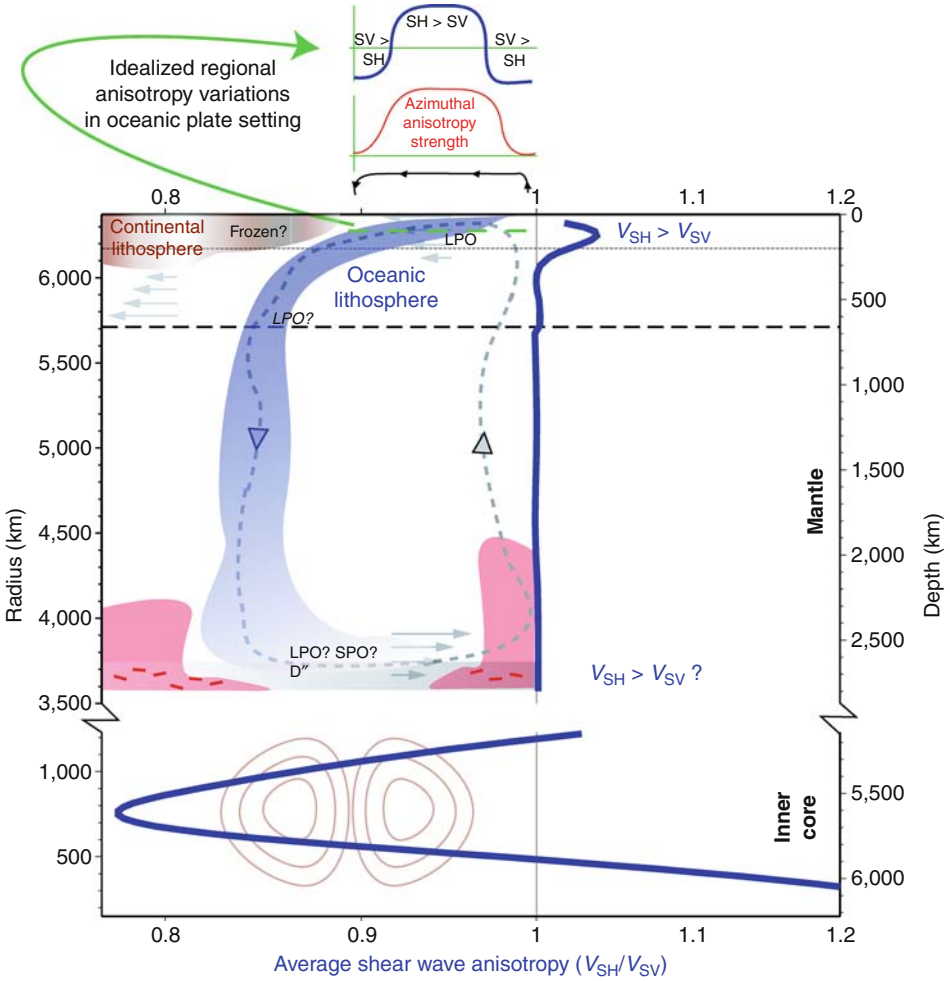
Thorsten W. Becker
Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA

Definition

Seismic anisotropy refers to the directional dependence of seismic wave speeds and particle motion polarizations, as well as the splitting of normal modes, as caused by the elastic properties of rocks.

Introduction

Many of the minerals that make up Earth are intrinsically anisotropic. When rocks are progressively deformed over geologic timescales, the alignment of mineral grains (lattice-preferred orientation, LPO) can lead to bulk anisotropy of the rock. Bulk anisotropy can additionally be generated by an ordered assembly of individually isotropic materials of different wave speeds (shape-preferred orientation, SPO). Both types of anisotropy are found within the Earth; SPO anisotropy also highlights a fundamental ambiguity between isotropic heterogeneity and anisotropy. Seismic wave propagation through an anisotropic medium depends on the wavelengths over which a particular wave type averages, complicating the analysis of seismological data. Both LPO and SPO imply significantly different (up to ~10%) speeds for waves of different polarization or propagation directions, and velocity variations can be larger than those expected from compositional or thermal heterogeneity. Seismic anisotropy is therefore of fundamental importance for structural imaging studies. To get robust estimates of the quantities of interest for geodynamic interpretation, the trade-off between isotropic and anisotropic structure has to be considered. Seismic anisotropy provides a powerful link between seismic observations and the dynamic processes that shape the solid Earth, for example, convective flow in the case of LPO in the mantle (Figure 1, see *Mantle Convection*). However, anisotropic tomographic inversions are inherently more nonunique than isotropic imaging because a general anisotropic, linearly elastic medium has 21 independent components of the elasticity tensor, as opposed to 2 in the isotropic case. As a consequence of the increased number of parameters and the differences in how data sampling constrains isotropic and anisotropic structure, more data are needed for the same level of resolution in an anisotropic inversion. Typically, additional a priori constraints, such as from



Seismic Anisotropy, Figure 1 Cartoon of the possible distribution of whole Earth anisotropy (note scale break at CMB) with geodynamic interpretation (cf. Montagner, 2007); *dotted* and *dashed horizontal lines* indicate 200 and 660 km depths, respectively. The *heavy blue lines* in center show average radial anisotropy from Kustowski et al. (2008) for the mantle and from Beghein and Trampert (2003) for the inner core. Underneath oceanic plates, mantle flow is primarily horizontal, leading to LPO anisotropy alignment with $v_{SH} > v_{SV}$, while the radial mass transport associated with upwellings and downwellings may lead locally $v_{SV} > v_{SH}$. Beneath continental regions, both frozen in anisotropy from past tectonic deformation and asthenospheric anisotropy from present-day convection may contribute. The *gray, dashed, circular line* in the mantle indicates an idealized flow trajectory for a downwelling slab (*blue*) displacing a thermochemical "pile" (*red*) at the core mantle boundary (cf. Garnero and McNamara, 2008). This deep flow may affect CMB dynamics and lead to LPO and/or SPO anisotropy (Modified from Long and Becker, 2010).

petrology, are needed to narrow the parameter space. These complexities make the study of anisotropy in a geodynamic context inherently multidisciplinary, involving seismology, mineral physics, rock mechanics, and geodynamic modeling.

Basic mathematical description

Seismic anisotropy arises when the linear elasticity tensor C that connects stress, σ, and strain, ε, tensors as

$$\sigma_{ij} = \sum_{k=1}^{3} \sum_{l=1}^{3} C_{ijkl} \varepsilon_{kl} \tag{1}$$

does not simplify to the isotropic form

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right), \tag{2}$$

where λ and μ are the first and second (shear modulus) Lamé parameters, δ the Kronecker delta ($\delta_{ij} = 1$ for $i = j$, and 0 else), and the indices run over the three axes of the coordinate system, x_1, x_2, and x_3. In general, C has 81 components, out of which 21 are independent, and the most complex (triclinic) form of anisotropy requires the specification of all those components (e.g., Hearmon, 1961; Nye, 1985; Anderson, 1989). A typical simplification is to assume hexagonal anisotropy, which should capture most aspects of olivine LPO-related anisotropy in the upper

mantle (Montagner and Anderson, 1989; Browaeys and Chevrot, 2004; Becker et al., 2006).

If C is expressed in the 6 × 6 Voigt matrix, c, notation where c_{mn} relates to C_{ijkl} as $m = \delta_{ij}i + (1 - \delta_{ij})(9 - i - j)$, and $n = \delta_{kl}k + (1 - \delta_{kl})(9 - k - 1)$, then the five Love (1927) parameters for the transversely isotropic medium that results from hexagonal symmetry correspond to

$$\begin{aligned} A &= c_{11} = \rho v_{\mathrm{PH}}^2, \quad C = c_{33} = \rho v_{\mathrm{PV}}^2, \\ L &= c_{44} = \rho v_{\mathrm{SV}}^2, \quad N = c_{66} = \rho v_{\mathrm{SH}}^2, \quad \text{and} \\ F &= c_{11}. \end{aligned} \tag{3}$$

Here, ρ is density, and $v_{\mathrm{PH,\,PV}}$ the velocities for P waves propagating horizontally (H, in x_1–x_2 plane) and vertically (V, x_3-axis), respectively. For shear waves, $v_{\mathrm{SH,SV}}$ in Eq. 3 are the velocities for horizontally propagating waves that are horizontally or vertically polarized, respectively (see *Propagation of Elastic Waves: Fundamentals*). Transverse isotropy as a simplified description of material anisotropy is widely used and developed in exploration seismics (e.g., Thomsen, 1986). The top 220 km in the PREM 1D Earth model (Dziewoński and Anderson, 1981) are also transversely isotropic with vertical symmetry axis as in Eq. 3; such a medium is said to have bulk radial anisotropy. (Note that vertically propagating S waves in this case have the same velocity, v_{SV}, regardless of polarization direction.)

Different combinations of the Love parameters or c_{nm} are used in the literature (e.g., Babuška and Cara, 1991); for example, anisotropy in PREM is described by two measures of shear- and compressional-wave anisotropy strength

$$\xi = \left(\frac{v_{\mathrm{SH}}}{v_{\mathrm{SV}}}\right)^2 = \frac{N}{L} \quad \text{and} \quad \varphi = \left(\frac{v_{\mathrm{PV}}}{v_{\mathrm{PH}}}\right)^2 = \frac{C}{A}, \tag{4}$$

respectively, and the parameter $\eta = F/(A - 2L)$, which controls how velocities change between the vertical and horizontal directions. Another way to characterize the anisotropy of a transversely isotropic medium is due to Thomsen (1986), who defined

$$\begin{aligned} \varepsilon &= \frac{c_{11} - c_{33}}{2c_{33}} = \frac{A - C}{C} \quad \text{and} \\ \gamma &= \frac{c_{66} - c_{44}}{2c_{44}} = \frac{N - L}{L} \end{aligned} \tag{5}$$

as two different measures of the P and S wave anisotropy strength, respectively, and a combined parameter

$$\delta^* = \frac{1}{2c_{33}^2}\left[2\left(c_{13} + c_{44}\right)^2 - (c_{33} - c_{44})(c_{11} + c_{33} - 2c_{44})\right], \tag{6}$$

which, for weak anisotropy, simplifies to

$$\delta = \frac{(c_{13} + c_{44})^2 - (c_{33} - c_{44})^2}{2c_{33}(c_{33} - c_{44})}. \tag{7}$$

The δ parameter is important for near-vertical P wave propagation and identical to ε for "elliptical" anisotropy (Thomsen, 1986). Mainprice (2007) provides an account of other combinations of c_{mn} in use to characterize a transversely isotropic medium. Those differ, regrettably, quite substantially in different fields of study.

If the symmetry axis of the hexagonal anisotropy is in the horizontal plane, the anisotropy is termed azimuthal. This means that perpendicular fast and slow axes can be defined for horizontally propagating SV waves, where waves will propagate with $v_{\mathrm{SV1}} > v_{\mathrm{SV2}}$ along the fast and slow orientations, respectively. Any perturbations to phase velocity p, δp, due to general, but small anisotropy can be expressed as a series of isotropic, π-periodic, and π/2 periodic terms (e.g., Backus, 1965; Forsyth, 1975):

$$\frac{\delta p}{p} \approx A_0 + A_1 cos(2\Psi) + A_2 sin(2\Psi) + A_3 cos(4\Psi) + A_4 sin(4\Psi). \tag{8}$$

Here, Ψ is the azimuth of wave propagation, and Eq. 8 follows from the wave equation and the rank of the elasticity tensor (Smith and Dahlen, 1973). For mantle rocks, the 2Ψ terms are expected to be larger than the 4Ψ contributions for Rayleigh waves, which are predominantly sensitive to SV (Anderson, 1966; Montagner and Nataf, 1986). The 4Ψ terms are expected to be bigger than 2Ψ for Love waves, motivating the focus on Rayleigh waves for azimuthal anisotropy studies (see *Surface Waves*).

In general, the wave propagation effects of any elasticity tensor C can be analyzed by considering a plane wave $\mathbf{u} = \mathbf{a}\exp(-i\omega(t - \mathbf{s}\cdot\mathbf{x}))$ with ω angular frequency, and **u**, **a**, **s**, and **x** the displacement, polarization, slowness, and location vectors, respectively (see *Propagation of Elastic Waves: Fundamentals*). **s** shall have the normalized direction $\check{\mathbf{s}}$ and length of $1/p$. Using the momentum equation $\ddot{u}_i = \partial_j \sigma_{ij}$, Eq. 1, the definition of the strain tensor, $\varepsilon_{ij} = \frac{1}{2}\left(\partial_i u_j + \partial_j u_i\right)$, C's symmetries, and defining $M_{ij} = \frac{1}{\rho} C_{ijkl}\,\hat{s}_j\,\hat{s}_l$, we can write

$$M\mathbf{a} = p^2\mathbf{a}, \tag{9}$$

which is an eigen problem for the symmetric matrix M. Equation 9 is called the Christoffel equation (e.g., Babuška and Cara, 1991). The eigen vector solutions correspond to one quasi-P and two quasi-S wave directions, and the associated eigen values are the density–velocity products ρv_{P}^2, ρv_{S2}^2, and ρv_{S1}^2. These quantities can be contoured for any elasticity tensor, for example as measured from single crystals, as a function of incidence angle and azimuth, to visualize the anisotropic properties of minerals or rocks in isolation (e.g., Mainprice, 2007). To generate more realistic synthetic body waves from three-dimensional (3D) variations in anisotropy,

semi-analytical reflectivity methods can be used if anisotropy is assumed to vary only with depth. However, for the general case of 3D variations of anisotropy on scales smaller than a Fresnel zone full, numerical wave propagation solutions are needed.

Seismological methods

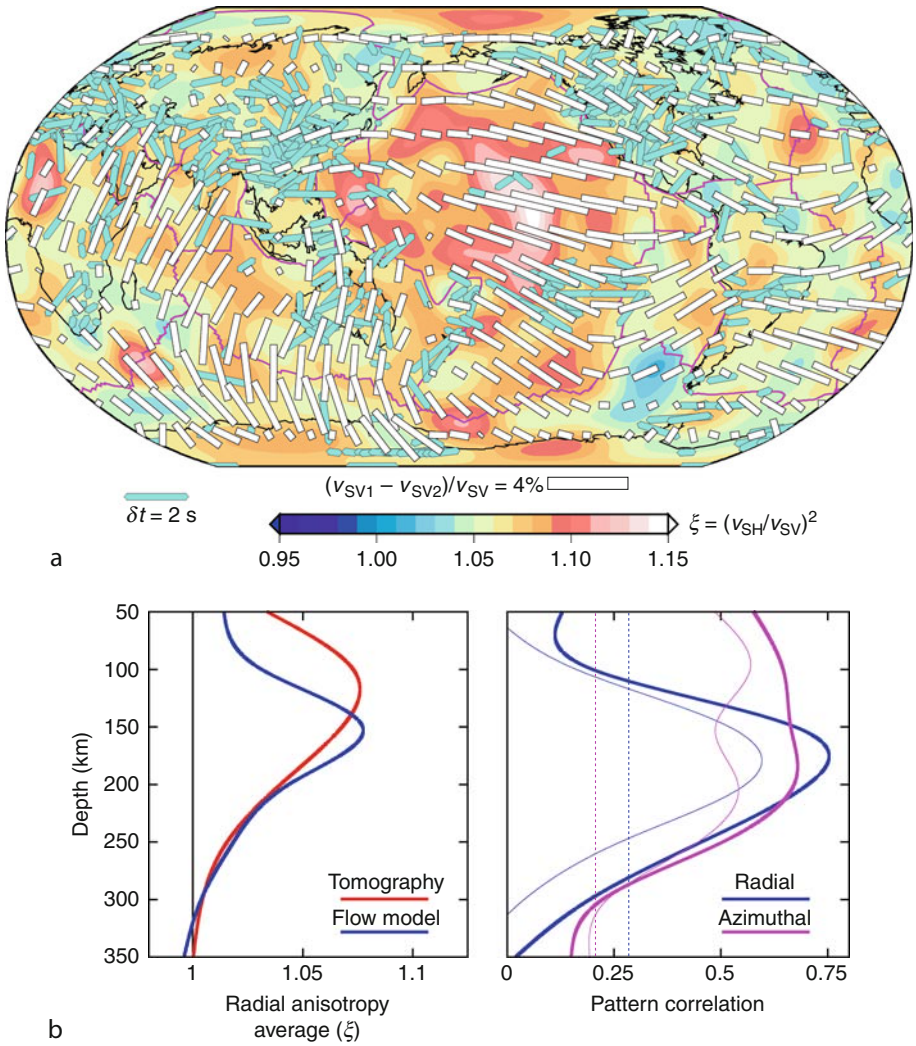
Seismic anisotropy can be detected in a number of ways that can be broadly classified into body- and surface-wave methods. The split of a shear wave into a fast and slow polarization direction as discussed for the solutions of Eq. 9 is akin to optical birefringence; it is exploited by the most popular method of measuring anisotropy, that utilizing *shear wave splitting* (Ando et al., 1983; Vinnik et al., 1984; Silver and Chan, 1991). For lithospheric and mantle applications, one typically considers near-vertical incidence SKS or SKKS core phases (see *Body Waves*), because the effects of any source-side anisotropy are removed by the S-to-P-to-S conversion upon traversal of the core. The most common splitting measurement consists of detecting the horizontal orientation of the fast (azimuth Ψ) pseudo-S wave from recorded particle motions, as well as determining the delay time δt between the arrival of the fast and slow *S* pulses (e.g., Savage, 1999; Long and Silver, 2009).

Shear wave splitting can be detected using a single earthquake measured at a single station, if wave propagation is out of a symmetry axis, and is a unique indicator for the presence of anisotropy along the ray path. However, only the highly idealized case of a single, transversely isotropic layer with horizontal symmetry axis can be directly interpreted in terms of Ψ and δt. Dipping symmetry axes, non-hexagonal anisotropy, or variations of anisotropy with depth will all cause a dependence of apparent splitting on back-azimuth (e.g., Schulte-Pelkum and Blackman, 2003). The nonlinear nature of the splitting measurement and layer splitting itself can lead to a bias of sensitivity toward the surface (∼one wavelength under the station), and not simple superposition (e.g., Saltzer et al., 2000). Such complexities make it imperative to strive for good back-azimuthal coverage, requiring the recording of several, suitable earthquakes, which is often a challenge given station-event geometry, or the duration of temporary deployments. If back-azimuth variations are detected, those can be used to make inferences about the variation of anisotropy with depth, which is undefined based on isolated measurements where anisotropy could, in principle, arise anywhere between the core mantle boundary (CMB) and the surface in the case of SKS splitting. If regional S arrivals are used, crossing ray paths can be used to infer 3D variations of anisotropy (e.g., Abt and Fischer, 2008). For teleseismic arrivals, the use of sensitivity kernels (e.g., Chevrot, 2006; Long et al., 2008) for the multichannel type of measurement of splitting holds great promise for resolving 3D anisotropy in regions for which close (closer than Fresnel zone width) station spacing is available. Broadly speaking, shear wave splitting is, however, a measurement with good lateral (∼50 km), but fairly poor depth resolution (Savage, 1999).

Another single, body-wave arrival method that follows from Eq. 9 is to use the orientation of the pseudo-*P* polarization, which may differ by more than 10° from along-ray, for *P* polarization anisotropy (Schulte-Pelkum et al., 2001), P_{pol}. A measurement of P_{pol} is sensitive to ∼half a wavelength underneath the station. If several, near-horizontal *P* paths with different azimuths are available, as in the case of the refracted P_n phase, which senses underneath the Moho, velocities can be plotted against azimuth to infer azimuthal anisotropy. This method was used for one of the earliest demonstrations of seismic anisotropy by Hess (1964), and a global comparison of P_n and SKS splitting can be found in Smith and Ekström (1999). The variations in delay times of teleseismic waves can also be used directly to infer anisotropy in the mantle lithosphere (e.g., Babuška et al., 1984; Bokelmann, 2002) by means of a tomographic inversion (see *Seismic Tomography*), but this method requires knowledge of the isotropic variations in wave speeds. This is a challenge for all tomographic approaches to anisotropy because there will always be a trade-off between isotropic and anisotropic heterogeneity in the absence of perfect data coverage (e.g., Tanimoto and Anderson, 1985). In terms of their depth-sensitivity, the body-wave methods can be ranked, from shallow to deep, as P_n, P_{pol}, SKS, and *P* delay times (Schulte-Pelkum and Blackman, 2003).

At crustal depths, anisotropy can additionally be detected by wide-angle refraction methods (e.g., Meissner et al., 2002, 2006). Orthogonal profiling, for example, may show a mismatch of derived seismic velocities, or a mismatch of refraction and reflection depths that can be interpreted in terms of anisotropic wave propagation. Receiver function methods (e.g., Park and Levin, 2002) (see *Seismic, Receiver Function Technique*) yield further evidence of crustal anisotropy from the existence of split *pS* conversions. Azimuthal variations in radial-transverse receiver function amplitudes are diagnostic of anisotropy versus tilted structure, and of the amount of anisotropy (e.g., Savage, 1998).

A wealth of information about anisotropy arises from the study of *surface waves*. The observation that Love waves, which mainly sense SH, travel faster than Rayleigh waves, which mainly sense SV due to their intrinsic polarities, implies the existence of a mean radial anisotropy in the upper mantle (Anderson, 1966; Dziewoński and Anderson, 1981). The existence of azimuthal anisotropy was documented for the Pacific by the study of Rayleigh waves (Forsyth, 1975), and Nataf et al. (1984) presented radially anisotropic, upper-mantle tomography. The current state of tomographic models for global azimuthal and radial anisotropy patterns is discussed by Montagner (2007). Surface wave propagation is dispersive, which allows the construction of 3D models of anisotropy (see *Earth's Structure, Global*; *Seismology, Global Earthquake Model*). The most easily measured phase-velocity period range for fundamental modes



Seismic Anisotropy, Figure 2 Global, uppermost mantle seismic anisotropy. (**a**) Seismological constraints: Radial (background, from Kustowski et al., 2008) and azimuthal anisotropy (white sticks indicating fast orientation, from Lebedev and van der Hilst, 2008) at 150 km, as well as SKS splitting (cyan sticks). SKS data are shown as a simple 5° average of the compilations by Fouch (2006) and Wüstefeld et al. (2009), but note that such averaging only provides a simplified view of azimuthal anisotropy (see text, and Schulte-Pelkum and Blackman, 2003, for example). (**b**) Radial anisotropy layer averages, on left, for the seismological model of Kustowski et al. (2008) and as predicted from the best-fitting geodynamic model of Becker et al. (2008). On right, pattern correlations up to spherical harmonic degree 8 between the same geodynamic model and radial (from Kustowski et al., 2008) and azimuthal (from Lebedev and van der Hilst, 2008) seismic tomography. *Heavy* and *thin lines* denote oceanic lithosphere only and global correlations, respectively. *Vertical, dashed lines* show 95% significance level (Modified from Long and Becker, 2010).

between ~50 and 150 s samples from the lithosphere down to ~300 km depth, and Figure 2a shows azimuthal and radial anisotropy at an asthenospheric depth of 150 km as well as a global compilation of SKS splitting results. At the shallow end, array methods (e.g., Deschamps et al., 2008) and in particular noise tomography (e.g., Lin et al., 2008) (see *Seismic, Ambient Noise Correlation*) facilitate extending the period range to shorter, crustal periods. Overtones can be used to constrain the deeper mantle, down to the 660-km phase transition (e.g., Trampert and van Heijst, 2002; Lebedev and van der Hilst, 2008). Lastly, the long-period surface wave equivalent of *free oscillations of the Earth* can provide constraints on the deep Earth, including inner core anisotropy (e.g., Woodhouse et al., 1986; Tromp, 2001).

Surface wave studies of anisotropy have fairly good depth sensitivity in that they are able to locate the origin of anisotropic signals in the upper mantle to within ~100 km depth. However, particularly compared to body wave measurements such as SKS splitting, the lateral resolution of surface waves is limited, for isotropic structure to perhaps ~500 and ~50 km for global and regional models, respectively, at present. Reasons for discrepancies between published tomographic models include the

different treatment of crustal corrections and phase-velocity measurements, as well as theoretical assumptions about wave propagation. Perhaps more important factors are the globally uneven ray path coverage or regularization choices.

A complete, 3D model of general seismic anisotropy would allow for more powerful petrological and geodynamic inferences than limited studies that focus only on a few aspects of anisotropy or wave propagation. Given the wide array of seismological observations, a desirable procedure to constrain the full 3D dependence of anisotropy is to compare different measures of anisotropy (e.g., Montagner et al., 2000; Wüstefeld et al., 2009) or to perform a joint inversion (e.g., Montagner and Nataf, 1988; Šíleny and Plomerová, 1996; Marone and Romanowicz, 2007). Sensitivity kernels that account for finite-frequency wave-propagation effects and the resulting complex 3D structure of a measurement's sensitivity to Earth structure (e.g., Chevrot, 2006; Long et al., 2008; Sieminski et al., 2009) can facilitate the relative weighting of different observations. Likewise, the incorporation of petrological constraints (e.g., Montagner and Anderson, 1989; Becker et al., 2006) can be used to simplify inversions further (Panning and Nolet, 2008; Chevrot and Monteiller, 2009).

Origin of anisotropy

The SPO type of anisotropy may be caused by any consistent alignment of entities with different isotropic wave speeds. Examples include lower crustal lamellae structures, cracks, or melt inclusions (e.g., Mainprice and Nicholas, 1989; Weiss et al., 1999; Meissner et al., 2006). Crack alignment will be only important for the shallowest crust where it may be indicative of crustal stress (e.g., Crampin and Chastin, 2003). Alignment of partial melt pockets may play a role both for shallow, extensional lithospheric domains, such as underneath mid-oceanic spreading centers or intracontinental rifts (e.g., Holtzman et al., 2003), and at the base of the mantle in the *Mantle D″ Layer* (e.g., Moore et al., 2004).

In between, the LPO type of anisotropy caused by the alignment of intrinsically anisotropic crystals is the most likely cause of anisotropy. The fundamental symmetry classes of material anisotropy of the constituent minerals (e.g., Nye, 1985; Anderson, 1989) determine the overall type of anisotropy in the Earth, and wave propagation depends critically on the type of anisotropy (e.g., Levin and Park, 1998; Schulte-Pelkum and Blackman, 2003). Several crustal rocks show LPO anisotropy; of particular interest are those rich in phyllosilicates (micas) in the upper-middle crust, and amphibole minerals in lower crust (e.g., Christensen and Mooney, 1995). In the upper mantle, the highly anisotropic olivine makes up ~60% of rocks (e.g., Mainprice, 2007). Laboratory experiments show that if multi-crystal olivine assemblages are deformed in the dislocation creep regime, crystals typically align such that the resulting fast propagation orientation rotates into the direction of shear, and many mantle xenoliths show corresponding LPO patterns (Mainprice, 2007; Karato et al., 2008).

This connection between rock deformation and seismic anisotropy allows an association of the patterns of azimuthal mantle anisotropy (e.g., Figure 2a) with *mantle convection* (e.g., McKenzie, 1979; Tanimoto and Anderson, 1984). A coarse approximation uses tectonic plate motion to imply deep flow direction, or, more realistically, flow can be calculated from global circulation models (Hager and O'Connell, 1981). The general association between mantle flow and anisotropy in terms of radial anisotropy is that flow in the upper boundary layer aligns olivine such that $v_{SH} > v_{SV}$ underneath oceanic plates due to a simple shear type of deformation (Figure 1). In regions of dominantly radial mass transport such as subduction zones and underneath spreading centers, $v_{SV} > v_{SH}$ (Chastel et al., 1993; Montagner, 2007). The radial and azimuthal anisotropy patterns shown in Figure 2a are broadly consistent with this expectation (Figure 2b), though there are also clear differences which are easier to constrain in regional studies (e.g., Gaherty et al., 1996). Complexities include variations azimuthal anisotropy orientations and amplitudes (e.g., Ekström and Dziewonski, 1998; Smith et al., 2004), and many of those patterns are accessible to geodynamic modeling, discussed below.

Given the importance of the details of the connection between seismology and geodynamics, several theoretical descriptions exist that predict microstructural LPO development given general deformation histories, as constrained by laboratory experiments (e.g., Kaminski and Ribe, 2001; Blackman, 2007). However, further laboratory constraints, for example on the reorientation of existing LPO fabrics under changing deformation regimes, are required to decide on the most appropriate treatment. Complex deformation histories are expected to lead to complex anisotropy. Yet, under monotonous deformation (e.g., by simple shear), olivine LPO is expected to saturate over finite strains of ~10. Amplitude variations compared to a single crystal may therefore be mainly due to orientation of the symmetry axis of the effective elastic tensor for an aggregate of crystals (cf. Karato et al., 2008).

Laboratory work over the last 10 years has further shown that the role of water content, deviatoric stress levels, and pressure can lead to significantly different LPO development from the typical, dry A-type fabrics that show the "fast axes along flow" alignment discussed above. For example, the high stress, high water content B-type fabric aligns the fast axes of olivine orthogonal to the direction of shear. Variations in water content have been used to explain some of the variability that is apparent in asthenospheric depth anisotropy, such as the decrease in azimuthal anisotropy strength across the Pacific from young to older seafloor, or the variability of orientations of SKS splitting in subduction zones (Mainprice, 2007; Karato et al., 2008).

LPO development under deformation of mantle rocks not only affects seismic properties, but also leads to thermal and mechanical anisotropy. The feedback of these effects into mantle convection and lithospheric deformation are potentially profound (e.g., Christensen, 1987; Chastel et al., 1993; Lev and Hager, 2008; Tommasi et al., 2009) and are currently an active area of research.

Observations of anisotropy and dynamic inferences

Whole earth anisotropy

Seismic anisotropy is found throughout the Earth, with the exception of the fluid outer core, though it is concentrated in certain depth regions (Figure 1). In the mantle, the best-constrained and strongest signal is found in the uppermost ~300 km where SH velocities are faster than SV by up to ~4% on average, as indicated by the Love–Rayleigh discrepancy. The exact shape of the average radial anisotropy profile is less certain, though most recent models agree that the largest anomalies are not found at the surface, but rather at ~100 km depth (Figure 2b). This peak may be associated with asthenospheric shear flow which is expected to lead to the largest strain-rates underneath the oceanic lithosphere, which is up to ~100 km thick when defined thermally (see *Mantle Convection*). Given that mantle anisotropy is most likely caused by LPO of olivine, the peak in seismic anisotropy in the uppermost mantle has been associated with the relatively high stress and low temperature depth region where dislocation dominates over diffusion creep (Karato, 1992; Gaherty and Jordan, 1995) (see *Mantle Viscosity*). Using composite rheologies, geodynamic models can be used to estimate the transition depths for the different creep laws, so delineating the region where LPO forms explicitly (e.g., McNamara et al., 2002; Podolefsky et al., 2004; Becker et al., 2008). Once rocks transition into the diffusion-creep dominated deformation regime, LPO is typically assumed to be destroyed quickly at high temperatures, or left preserved (frozen in) at low temperatures/small velocity gradients. The decrease in radial anisotropy toward the surface (Figure 2b) may therefore be associated with tectonically older, frozen in structure. On the scales accessible by surface wave studies, for example, anisotropy in old lithospheric domains may be less well aligned into the vertical, or into a coherent horizontal orientation, than in the asthenosphere, which is shaped by current mantle convection (e.g., Fouch and Rondenay, 2006).

At larger mantle depths, radial anisotropy becomes less well constrained (e.g., Visser et al., 2008). There is some indication that radial anomalies pick up around the transition zone (Figure 1), and several studies have argued for the existence of azimuthal anisotropy around 660 km (e.g., Trampert and van Heijst, 2002; Wookey et al., 2002). Most of the lower mantle is nearly isotropic until the D″ region close to the core mantle boundary where there is good evidence for the existence of anisotropy from regional studies (e.g., Moore et al., 2004), and indications for average radial anisotropy from global studies (Boschi and Dziewoński, 2000; Panning and Romanowicz, 2006). As for the upper mantle, one may invoke an LPO reactivation of dislocation creep, for example in cold, highly deformed subduction slabs (see Figure 1; McNamara et al., 2002). The other, at present perhaps equally likely, mechanism that has been invoked for D″ anisotropy is the alignment of melt tubules (SPO). Melt alignment may also play a role in the transition zone if the latter represents a melt-rich water filter (Bercovici and Karato, 2003). The D″ region is expected to be at least as dynamically complex as the upper thermal boundary layer, and both domains are affected by compositional anomalies. Those include the continental lithosphere, with its stiff, compositionally anomalous and presumably neutrally buoyant cratonic keels, and likely piles of dense material at the base of the mantle in regions displaced along the CMB from recent subduction (e.g., Garnero, 2004; Garnero and McNamara, 2008). We therefore expect significant lateral variations in the generation of anisotropy within D″ depending on the vertical flow setting (Figure 1, e.g., Moore et al., 2004). Close to the CMB, anisotropy may also vary with depth depending on if lower mantle material has transitioned to the post-perovskite phase (e.g., Wookey et al., 2005; Merkel et al., 2007).

There is also robust evidence for anisotropy within the Earth's core. Body waves that traverse the inner core and are aligned with the rotation axis arrive earlier than those that cross in the equatorial plane (Morelli et al., 1986). Evidence for anisotropy is also seen in the splitting of normal modes (Woodhouse et al., 1986), and more recent data and models for core anisotropy are discussed in Tromp (2001) and Souriau (2007). However, there are still debates on the exact nature of the anisotropy distribution with depth (cf. Ishii and Dziewoński, 2003). Figure 1 shows radial, shear-wave anisotropy for the inner core from Beghein and Trampert (2003). This particular model invoked a hexagonal close-packed phase of iron in the upper half of the inner core, and perhaps a transition into a different iron phase at depth, and predicts large amplitudes of radial anisotropy compared to the upper mantle. The origin of inner core anisotropy is also less clear than for the upper mantle (Mainprice, 2007). One hypothesis that has recently been discussed in some detail is freezing in of convective patterns during the cooling and evolution of the inner core (Jeanloz and Wenk, 1988; Buffett, 2009; Deguen and Cardin, 2009).

Structure and dynamics of the upper boundary layer

Seismic anisotropy at every depth range throughout the Earth holds valuable information on the dynamics of the planet. The connections can be made quantitative most easily for the shallower layers where seismological constraints abound, rock deformation is accessible via laboratory experiments, and geodynamic modeling is fairly well constrained. In the case of crack anisotropy in the shallow

crust, observations yield constraints on regional stress fields. Applications include industry work (vertical seismic profiling in boreholes), earthquake studies around faults, and volcano monitoring where cracking due to magma migration can be traced.

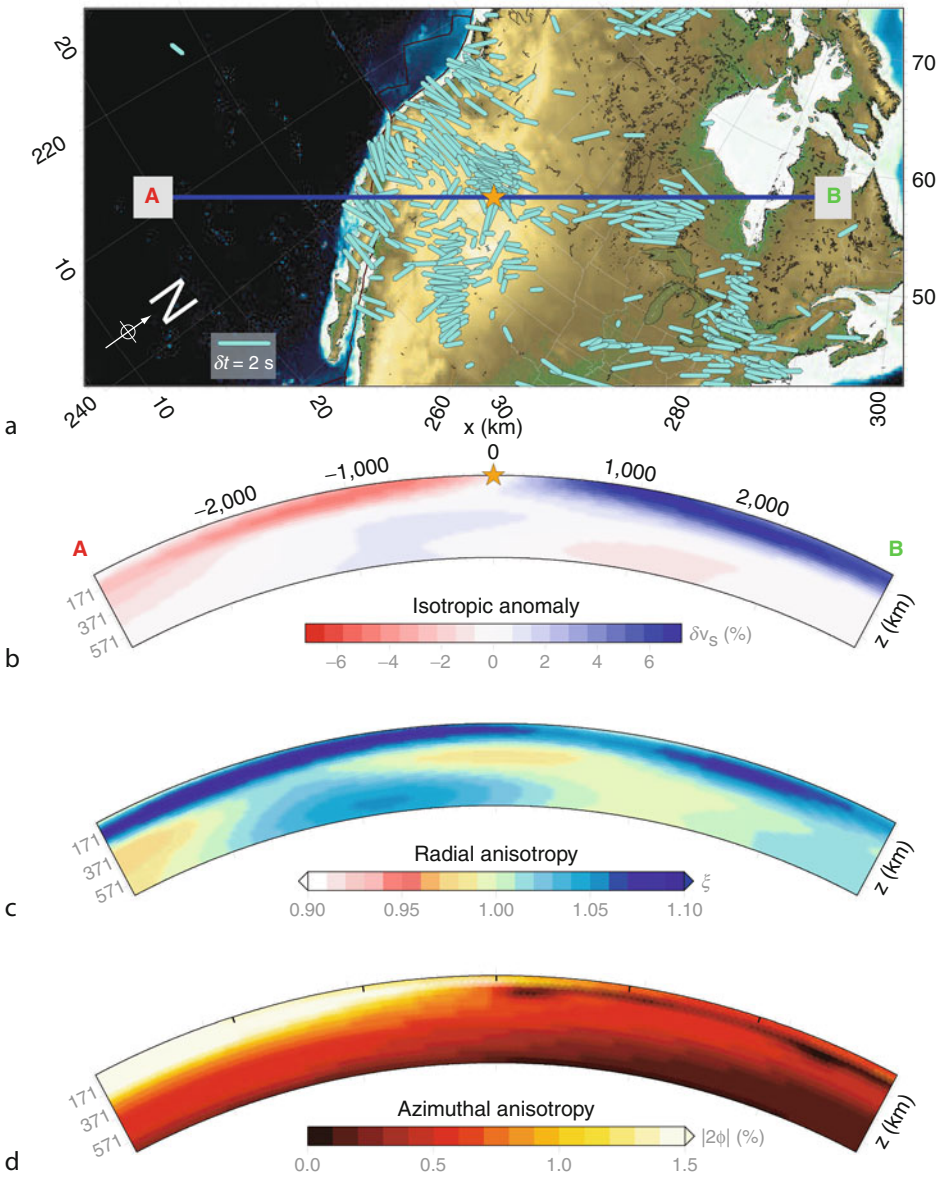
Within the upper convective boundary layer, the oceanic plate domains (see *Lithosphere, Oceanic*) should most closely resemble the simplified view of radial and azimuthal anisotropy due to LPO anisotropy formation in mantle flow as shown in Figure 1. Gaboret et al. (2003), Becker et al. (2003), and Behn et al. (2004) showed that mantle circulation from geodynamic models does indeed provide a valid explanation for azimuthal anisotropy patterns (Figure 2a), and that comparison of model predictions with anisotropy can yield constraints on mantle flow, such as the role of buoyant mantle upwellings as opposed to plate-induced shear. Becker et al. (2008) provided a quantitative model of radial anisotropy, and Figure 2b shows the fit of their preferred model to radial anisotropy averages in the upper mantle, as well as lateral patterns in azimuthal and radial anisotropy. Results are consistent with the expectation that the geodynamic models should describe recent (few 10 s of Myr) asthenospheric flow best. The correlations between geodynamics and the seismological models (Figure 2b) are comparable or better than the match between different seismological models. Such first-order agreement between global geodynamics and seismology motivates current modeling efforts, for example on constraining the amount of net rotations of the lithosphere or the degree of lateral viscosity variations (e.g., Becker, 2008; Conrad et al., 2007; Conrad and Behn, 2010; Kreemer, 2009).

Figure 2b shows that geodynamic models typically underpredict radial anisotropy in the shallower parts of the lithosphere, which is mainly due to continental domains. While anisotropy in younger continental lithosphere such as in the western USA appears to be well described by asthenospheric flow, older regions show more complex behavior such as a consistent orientation of seismic anisotropy over several hundred kilometers (e.g., Babuška and Plomerová, 2006). It has been suggested that anisotropy is concentrated in, and frozen into, the continental lithosphere, or, alternatively, that radial anisotropy is largest right underneath the mechanical boundary layer formed by stiff continents (e.g., Gaherty and Jordan, 1995; Gung et al., 2003; Fouch and Rondenay, 2006). Figure 3 shows a profile through North America; anisotropy, as inferred from these models, only partially conforms to the simplified expectations (cf. Panning and Romanowicz, 2006). The cross section of radial anisotropy shows the expected focusing of SH faster than SV in the Pacific plate, and some regionally enhanced $v_{SH} > v_{SV}$ within the eastern USA and the Canadian craton, but no enhanced anisotropy beneath what would be inferred to be the base of the continental lithosphere from the isotropic anomalies. Azimuthal anisotropy is also, expectedly, strong within the Pacific plate (compare Figure 2a), but there is an intriguing low-azimuthal anisotropy channel within the eastern North American continental lithosphere. If such features are due to complex tectonic deformation with small lateral shear-coherence, or due to the averaging properties of surface waves and incomplete ray illumination, remains to be determined. The study of continental anisotropy is an active area of research, and many questions such as to the vertical coherence of lithospheric deformation and the depth extent of fault zone localization will benefit from the information that seismic anisotropy can bring to the table. There are numerous other, regional tectonic settings where anisotropy can yield important constraints, and those cannot be comprehensively reviewed here. Important examples include continental transforms and collision zones, spreading centers, and subduction zones. Reviews of our current understanding of such settings can be found in Silver (1996); Savage (1999); Park and Levin (2002) and Long and Silver (2009).

Powerful dynamic insights notwithstanding, there are still large uncertainties in every step of the chain of modeling that has to be followed. Complexities arise from inferring mantle flow from geodynamics (e.g., role of chemical vs. thermal buoyancy, uncertainties about rheology), to predicting LPO textures (e.g., proper microphysical treatment), to inferring elasticity tensors (e.g., homogenization and temperature/pressure derivatives), to mapping those tensors in 3D to whatever seismological observable (preferred) or seismological model (more common) are used to benchmark the models (e.g., finite frequency wave propagation, sampling). The finding that overall patterns appear to be well explained (Figure 2), and that synthetic LPOs match those of xenolith samples provide some *a posterior* justification for the modeling rationale. Moreover, these agreements indicate that the bulk of the asthenospheric flow is indeed dominated by dry A-type fabrics. However, future refinements of seismological imaging, for example through array deployments such as EarthScope USArray and temporary seafloor studies, theoretical developments in seismology, and the improved geodynamic treatment of anisotropy will undoubtedly lead to adjustment of our understanding of whole Earth anisotropic structure.

Summary

Seismic anisotropy is ubiquitous throughout the Earth and provides constraints on dynamic processes, from the stress in the crust, the origin and evolution of the continental lithosphere, through convective flow in the upper mantle, to core evolution. The state of upper-mantle geodynamic modeling is such that important questions, such as about absolute plate motion reference frames, intraplate deformation, or the hydration state of the mantle can be addressed. Important issues about the resolution of different seismological datasets and degree of robustness of seismological images remain. Joint with the inherent uncertainties in geodynamic modeling and how to map flow into seismic anisotropy, this means that numerous



Seismic Anisotropy, Figure 3 Pacific and North American upper mantle anisotropy. (**a**) SKS splitting (as in Figure 2, but averaged by 0.5°) and location of cross-continental profile; (**b**) isotropic shear wave velocity relative to background (Voigt average, from Kustowski et al., 2008); (**c**) radial anisotropy ($\xi = (v_{SH}/v_{SV})^2$, from Kustowski et al., 2008); and (**d**) strength of azimuthal anisotropy ($|(v_{SV1} - v_{SV2})/v_{SV}|$, from Lebedev and van der Hilst, 2008). (Modified from Long and Becker, 2010.)

questions for the interpretation of anisotropy observable are open. This challenge mandates further theoretical and instrumentational efforts and that the study of anisotropy proceeds interdisciplinary and in a dynamics context. Answering those questions holds the promise of arriving at a new understanding of the workings of the mantle system.

Acknowledgments

Detailed comments by Donna Blackman, Mark Behn, and Sergei Lebedev and valuable suggestions from Lapo Boschi, Sebastien Chevrot, David Okaya, Mark Panning, Vera Schulte-Pelkum, and an anonymous reviewer helped improve this contribution.

Bibliography

Abt, D. L., and Fischer, K. M., 2008. Resolving three-dimensional anisotropic structure with shear-wave splitting tomography. *Geophysical Journal International*, **173**, 859–886.

Anderson, D. L., 1966. Recent evidence concerning the structure and composition of the Earth's mantle. In *Physics and Chemistry of the Earth*. Oxford: Pergmanon, Vol. 6, pp. 1–131.

Anderson, D. L., 1989. *Theory of the Earth*. Boston, MA: Blackwell Scientific Publications (Available online at http://caltechbook.library.caltech.edu/14/1/TheoryoftheEarth.pdf, accessed 01/2010).
Ando, M., Ishikawa, Y., and Yamasaki, F., 1983. Shear-wave polarization anisotropy in the mantle beneath Honshu, Japan. *Journal of Geophysical Research*, **88**, 5850–5864.
Babuška, V., and Cara, M., 1991. *Seismic Anisotropy in the Earth*. Dordrecht: Kluwer Academic Publishers.
Babuška, V., and Plomerová, J., 2006. European mantle lithosphere assembled from rigid microplates with inherited seismic anisotropy. *Physics of the Earth and Planetary Interiors*, **158**, 264–280.
Babuška, V., Plomerová, J., and Šíleny, J., 1984. Spatial variations of Presiduals and deep structure of the European lithospere. *Geophysical Journal of the Royal Astronomical Society*, **79**, 363–383.
Backus, G. E., 1965. Possible forms of seismic anisotropy of the uppermost mantle under oceans. *Journal of Geophysical Research*, **70**, 3429–3439.
Becker, T. W., 2008. Azimuthal seismic anisotropy constrains net rotation of the lithosphere. *Geophysical Research Letters*, **35**, L05303, doi:10.1029/2007GL032928, correction: doi:10.1029/2008GL033946.
Becker, T. W., Kellogg, J. B., Ekström, G., and O'Connell, R. J., 2003. Comparison of azimuthal seismic anisotropy from surface waves and finite-strain from global mantle-circulation models. *Geophysical Journal International*, **155**, 696–714.
Becker, T. W., Chevrot, S., Schulte-Pelkum, V., and Blackman, D. K., 2006. Statistical properties of seismic anisotropy predicted by upper mantle geodynamic models. *Journal of Geophysical Research*, **111**, B08309, doi:10.1029/2005JB004095.
Becker, T. W., Kustowski, B., and Ekström, G., 2008. Radial seismic anisotropy as a constraint for upper mantle rheology. *Earth and Planetary Science Letters*, **267**, 213–237.
Beghein, C., and Trampert, J., 2003. Robust normal mode constraints on inner-core anisotropy from model space search. *Science*, **299**, 552–555.
Behn, M. D., Conrad, C. P., and Silver, P. G., 2004. Detection of upper mantle flow associated with the African Superplume. *Earth and Planetary Science Letters*, **224**, 259–274.
Bercovici, D., and Karato, S-i., 2003. Whole-mantle convection and the transition-zone water filter. *Nature*, **425**, 39–44.
Blackman, D., 2007. Use of mineral physics, with geodynamic modelling and seismology, to investigate flow in the Earth's mantle. *Reports on Progress in Physics*, **70**, 659–689.
Bokelmann, G. H. R., 2002. Convection-driven motion of the north American craton: evidence from *P*-wave anisotropy. *Geophysical Journal International*, **148**, 278–287.
Boschi, L., and Dziewoński, A. M., 2000. Whole Earth tomography from delay times of P, PcP, PKP phases: lateral heterogeneities in the outer core, or radial anisotropy in the mantle? *Journal of Geophysical Research*, **105**, 25,567–25,594.
Browaeys, J., and Chevrot, S., 2004. Decomposition of the elastic tensor and geophysical applications. *Geophysical Journal International*, **159**, 667–678.
Buffett, B., 2009. Onset and orientation of convection in the inner core. *Geophysical Journal International*, **179**, 711–719.
Chastel, Y. B., Dawson, P. R., Wenk, H.-R., and Bennett, K., 1993. Anisotropic convection with implications for the upper mantle. *Journal of Geophysical Research*, **98**, 17,757–17,771.
Chevrot, S., 2006. Finite-frequency vectorial tomography: a new method for high-resolution imaging of upper mantle anisotropy. *Geophysical Journal International*, **165**, 641–657.
Chevrot, S., and Monteiller, V., 2009. Principles of vectorial tomography – the effects of model parametrization and regularization in tomographic imaging of seismic anisotropy. *Geophysical Journal International*, **179**, 1726–1736.
Christensen, N. I., and Mooney, W. D., 1995. Seismic velocity structure and composition of the continental crust: a global review. *Journal of Geophysical Research*, **100**, 9761–9788.
Christensen, U. R., 1987. Some geodynamical effects of anisotropic viscosity. *Geophysical Journal of the Royal Astronomical Society*, **91**, 711–736.
Conrad, C. P., and Behn, M., 2010. Constraints on lithosphere net rotation and asthenospheric viscosity from global mantle flow models and seismic anisotropy. *Geochemistry, Geophysics, Geosystems*, **11**, Q05W05, doi:10.1029/2009GC002970.
Conrad, C. P., Behn, M. D., and Silver, P. G., 2007. Global mantle flow and the development of seismic anisotropy: differences between the oceanic and continental upper mantle. *Journal of Geophysical Research*, **112**, B07317, doi:10.1029/2006JB004608.
Crampin, S., and Chastin, S., 2003. A review of shear wave splitting in the crack-critical crust. *Geophysical Journal International*, **155**, 221–240.
Deguen, R., and Cardin, P., 2009. Tectonic history of the Earth's inner core preserved in its seismic structure. *Nature Geoscience*, **2**, 419–422.
Deschamps, F., Lebedev, S., Meier, T., and Trampert, J., 2008. Azimuthal anisotropy of Rayleigh-wave phase velocities in the east-central United States. *Geophysical Journal International*, **173**, 827–843.
Dziewoński, A. M., and Anderson, D. L., 1981. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, **25**, 297–356.
Ekström, G., and Dziewonski, A. M., 1998. The unique anisotropy of the Pacific upper mantle. *Nature*, **394**, 168–172.
Forsyth, D. W., 1975. The early structural evolution and anisotropy of the oceanic upper mantle. *Geophysical Journal of the Royal Astronomical Society*, **43**, 103–162.
Fouch, M., 2006. Upper mantle anisotropy database. Online, accessed in 06/2006, http://geophysics.asu.edu/anisotropy/upper/.
Fouch, M. J., and Rondenay, S., 2006. Seismic anisotropy beneath stable continental interiors. *Physics of the Earth and Planetary Interiors*, **158**, 292–320.
Gaboret, C., Forte, A. M., and Montagner, J.-P., 2003. The unique dynamics of the Pacific Hemisphere mantle and its signature on seismic anisotropy. *Earth and Planetary Science Letters*, **208**, 219–233.
Gaherty, J. B., and Jordan, T. H., 1995. Lehmann discontinuity as the base of an anisotropic layer beneath continents. *Science*, **268**, 1468–1471.
Gaherty, J. B., Jordan, T. H., and Gee, L. S., 1996. Seismic structure of the upper mantle in a central Pacific corridor. *Journal of Geophysical Research*, **101**, 22,291–22,310.
Garnero, E. J., 2004. A new paradigm for Earth's core-mantle boundary. *Science*, **304**, 835–836.
Garnero, E. J., and McNamara, A. K., 2008. Structure and dynamics of the Earth's lower mantle. *Science*, **320**, 626–628.
Gung, Y., Panning, M., and Romanowicz, B., 2003. Global anisotropy and the thickness of continents. *Nature*, **422**, 707–711.
Hager, B. H., and O'Connell, R. J., 1981. A simple global model of plate dynamics and mantle convection. *Journal of Geophysical Research*, **86**, 4843–4867.
Hearmon, R. F. S., 1961. *An Introduction to Applied Anisotropic Elasticity*. London: Oxford University Press.
Hess, H. H., 1964. Seismic anisotropy of the uppermost mantle under oceans. *Nature*, **203**, 629–631.
Holtzman, B. K., Kohlstedt, D. L., Zimmerman, M. E., Heidelbach, F., Hiraga, T., and Hustoft, J., 2003. Melt segregation and strain partitioning: implications for seismic anisotropy and mantle flow. *Science*, **301**, 1227–1230.

Ishii, M., and Dziewoński, A. M., 2003. Distinct seismic anisotropy at the centre of the earth. *Physics of the Earth and Planetary Interiors*, **140**, 203–217.

Jeanloz, R., and Wenk, H. R., 1988. Convection and anisotropy of the inner core. *Geophysical Research Letters*, **15**, 72–75.

Kaminski, É., and Ribe, N. M., 2001. A kinematic model for recrystallization and texture development in olivine polycrystals. *Earth and Planetary Science Letters*, **189**, 253–267.

Karato, S-i., 1992. On the Lehmann discontinuity. *Geophysical Research Letters*, **51**, 2255–2258.

Karato, S-i., Jung, H., Katayama, I., and Skemer, P., 2008. Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies. *Annual Review of Earth Planetary Sciences*, **36**, 59–95.

Kreemer, C., 2009. Absolute plate motions constrained by shear wave splitting orientations with implications for hot spot motions and mantle flow. *Journal of Geophysical Research,* **114**, B10405, doi:10.1029/2009JB006416.

Kustowski, B., Ekström, G., and Dziewoński, A. M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model. *Journal of Geophysical Research,* **113**, B06306, doi:10.1029/2007JB005169.

Lebedev, S., and van der Hilst, R. D., 2008. Global upper-mantle tomography with the automated multimode inversion of surface and S-wave forms. *Geophysical Journal International*, **173**, 505–518.

Lev, E., and Hager, B. H., 2008. Rayleigh Taylor instabilities with anisotropic lithospheric viscosity. *Geophysical Journal International*, **173**, 806–814.

Levin, V., and Park, J., 1998. *P – SH* conversions in layered media with hexagonally symmetric anisotropy: a cookbook. *Pure and Applied Geophysics*, **151**, 669–697.

Lin, F., Moschetti, M. P., and Ritzwoller, M. H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. *Geophysical Journal International*, **173**, 281–298.

Long, M. D., and Becker, T. W., 2010. Mantle dynamics and seismic anisotropy. *Earth and Planetary Science Letters,* **297**, 341–354.

Long, M. D., and Silver, P. G., 2009. Shear wave splitting and mantle anisotropy: measurements, interpretations, and new directions. *Surveys in Geophysics*, **30**, 407–461.

Long, M. D., de Hoop, M. V., and van der Hilst, R. D., 2008. Wave equation shear wave splitting tomography. *Geophysical Journal International*, **172**, 311–330.

Love, A. E. H., 1927. *A Treatise on the Mathematical Theory of Elasticity.* Cambridge: Cambridge University Press. reprinted in 1944 by Dover Publications, New York.

Mainprice, D., 2007. Seismic anisotropy of the deep Earth from a mineral and rock physics perspective. In Schubert, G., and Bercovic, D. (eds.), *Treatise on Geophysics*. Oxford: Elsevier, Vol. 2, pp. 437–492.

Mainprice, D., and Nicholas, A., 1989. Development of shape and lattice preferred orientations: application to the seismic anisotropy of the lower crust. *Journal of Structural Geology*, **11**, 175–189.

Marone, F., and Romanowicz, F., 2007. The depth distribution of azimuthal anisotropy in the continental upper mantle. *Nature*, **447**, 198–201.

McKenzie, D. P., 1979. Finite deformation during fluid flow. *Geophysical Journal of the Royal AstronomicalSociety*, **58**, 689–715.

McNamara, A. K., van Keken, P. E., and Karato, S-i., 2002. Development of anisotropic structure in the Earth's lower mantle by solid-state convection. *Nature*, **416**, 310–314.

Meissner, R., Mooney, W. D., and Artemieva, I., 2002. Seismic anisotropy and mantle creep in young orogens. *Geophysical Journal International*, **149**, 1–14.

Meissner, R., Rabbel, W., and Kern, H., 2006. Seismic lamination and anisotropy of the lower continental crust. *Tectonophysics*, **416**, 81–99.

Merkel, S., McNamara, A. K., Kubo, A., Speziale, S., Miyagi, L., Meng, Y., Duffy, T. S., and Wenk, H.-R., 2007. Deformation of (Mg, Fe)SiO3 post-perovskite and D″ anisotropy. *Science*, **316**(5832), 1729–32.

Montagner, J.-P., 2007. Upper mantle structure: Global isotropic and anisotropic elastic tomography. In Schubert, G., and Bercovici, D. (eds.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 1, pp. 559–589.

Montagner, J. P., and Anderson, D. L., 1989. Petrological constraints on seismic anisotropy. *Physics of the Earth and Planetary Interiors*, **54**, 82–105.

Montagner, J.-P., and Nataf, H.-C., 1986. A simple method for inverting the azimuthal anisotropy of surface waves. *Journal of Geophysical Research*, **91**, 511–520.

Montagner, J. P., and Nataf, H. C., 1988. Vectorial tomography-I. Theory. *Geophysical Journal*, **94**, 295–307.

Montagner, J.-P., Griot-Pommera, D.-A., and Laveé, J., 2000. How to relate body wave and surface wave anisotropy? *Journal of Geophysical Research*, **105**, 19,015–19,027.

Moore, M. M., Garnero, E. J., Lay, T., and Williams, Q., 2004. Shear wave splitting and waveform complexity for lowermost mantle structures with low-velocity lamellae and transverse isotropy. *Journal of Geophysical Research*, **103**, B02319, doi:10.1029/2003JB002546.

Morelli, A., Dziewonski, A. M., and Woodhouse, J. H., 1986. Anisotropy of the inner core inferred from PKIKP travel times. *Geophysical Research Letters*, **13**, 1545–1548.

Nataf, H.-C., Nakanishi, I., and Anderson, D. L., 1984. Anisotropy and shear velocity heterogeneity in the upper mantle. *Geophysical Research Letters*, **11**, 109–112.

Nye, J. F., 1985. *Physical Properties of Crystals*. London: Oxford University Press.

Panning, M., and Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophysical Journal International*, **167**, 361–379.

Panning, M. P., and Nolet, G., 2008. Surface wave tomography for azimuthal anisotropy in a strongly reduced parameter space. *Geophysical Journal International*, **174**, 629–648.

Park, J., and Levin, V., 2002. Seismic anisotropy: tracing plate dynamics in the mantle. *Science*, **296**, 485–489.

Podolefsky, N. S., Zhong, S., and McNamara, A. K., 2004. The anisotropic and rheological structure of the oceanic upper mantle from a simple model of plate shear. *Geophysical Journal International*, **158**, 287–296.

Saltzer, R. L., Gaherty, J. B., and Jordan, T. H., 2000. How are vertical shear wave splitting measurements affected by variations in the orientation of azimuthal anisotropy with depth? *Geophysical Journal International*, **141**, 374–390.

Savage, M., 1998. Lower crustal anisotropy or dipping boundaries? Effects on receiver functions and a case study in New Zealand. *Journal of Geophysical Research*, **103**, 15,069–15,087.

Savage, M. K., 1999. Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting? *Reviews of Geophysics*, **37**, 65–106.

Schulte-Pelkum, V., and Blackman, D. K., 2003. A synthesis of seismic *P* and *S* anisotropy. *Geophysical Journal International*, **154**, 166–178.

Schulte-Pelkum, V., Masters, G., and Shearer, P. M., 2001. Upper mantle anistropy from long-period *P* polarization. *Journal of Geophysical Research*, **106**, 21,917–21,934.

Sieminski, A., Trampert, J., and Tromp, J., 2009. Principal component analysis of anisotropic finite-frequency sensitivity kernels. *Geophysical Journal International*, **179**, 1186–1198.

Šíleny, J., and Plomerová, J., 1996. Inversion of shear-wave splitting parameters to retrieve three-dimensional orientation of anisotropy in continental lithosphere. *Physics of the Earth and Planetary Interiors*, **95**, 277–292.
Silver, P. G., 1996. Seismic anisotropy beneath the continents: probing the depths of geology. *Annual Review of Earth Planetary Sciences*, **24**, 385–432.
Silver, P. G., and Chan, W. W., 1991. Shear wave splitting and subcontinental mantle deformation. *Journal of Geophysical Research*, **96**, 16,429–16,454.
Smith, D. B., Ritzwoller, M. H., and Shapiro, N. M., 2004. Stratification of anisotropy in the Pacific upper mantle. *Journal of Geophysical Research*, **109**, B11309, doi:10.1029/2004JB003200.
Smith, G. P., and Ekström, G., 1999. A global study of P_n anisotropy beneath continents. *Journal of Geophysical Research*, **104**, 963–980.
Smith, M. L., and Dahlen, F. A., 1973. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. *Journal of Geophysical Research*, **78**, 3321–3333.
Souriau, A., 2007. The Earth's core. In Romanowicz, B., and Dziewonski, A. (eds.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 1, pp. 655–693.
Tanimoto, T., and Anderson, D. L., 1984. Mapping convection in the mantle. *Geophysical Research Letters*, **11**, 287–290.
Tanimoto, T., and Anderson, D. L., 1985. Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 s. *Journal of Geophysical Research*, **90**, 1842–1858.
Thomsen, L., 1986. Weak elastic anisotropy. *Geophysics*, **51**, 1954–1966.
Tommasi, A., Knoll, M., Vauchez, A., Sgnorelli, J., Thoraval, C., and Loge, R., 2009. Structural reactivation in plate tectonics controlled by olivine crystals anisotropy. *Nature Geosciences*, **2**, 423–427.
Trampert, J., and van Heijst, H. J., 2002. Global azimuthal anisotropy in the transition zone. *Science*, **296**, 1297–1299.
Tromp, J., 2001. Inner core anisotropy and rotation. *Annual Review of Earth and Planetary Sciences*, **29**, 47–69.
Vinnik, L., Kosarev, G. L., and Makeyeva, L. I., 1984. Anisotropy of the lithosphere from the observations of *SKS* and *SKKS* phases. *Proceedings of Academic Sciences USSR*, **278**, 1335–1339.
Visser, K., Trampert, J., Lebedev, S., and Kennett, B. L. N., 2008. Probability of radial anisotropy in the deep mantle. *Earth and Planetary Science Letters*, **270**, 241–250.
Weiss, T., Siegesmund, S., Rabbel, W., Bohlen, T., and Pohl, M., 1999. Seismic velocities and anisotropy of the lower continental crust: a review. *Pure and Applied Geophysics*, **156**, 97–122.
Woodhouse, J. H., Giardini, D., and Li, X.-D., 1986. Evidence for inner core anisotropy from free oscillations. *Geophysical Research Letters*, **13**, 1549–1552.
Wookey, J., Kendall, J.-M., and Barruol, G., 2002. Mid-mantle deformation inferred from seismic anisotropy. *Nature*, **415**, 777–780.
Wookey, J., Stackhouse, S., Kendall, J.-M., Brodholt, J., and Price, G. D., 2005. Efficacy of the post-perovskite phase as an explanation for lowermost-mantle seismic properties. *Nature*, **438**, 1004–1007.
Wüstefeld, A., Bokelmann, G. H. R., Barruol, G., and Montagner, J. P., 2009. Identifying global seismic anisotropy patterns by correlating shear-wave splitting and surface-wave data. *Physics of the Earth and Planetary Interiors*, **176**, 198–212.

Cross-references

Body Waves
Earth's Structure, Global
Free Oscillations of the Earth
Inverse Theory, Global Optimization
Inverse Theory, Linear
Inverse Theory, Monte Carlo Method
Inverse Theory, Singular Value Decomposition
Lithosphere, Continental
Lithosphere, Oceanic
Mantle Convection
Mantle D″ Layer
Mantle Viscosity
Propagation of Elastic Waves: Fundamentals
Seismic Tomography
Seismic, Ambient Noise Correlation
Seismic, Receiver Function Technique
Seismology, Global Earthquake Model
Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting
Surface Waves

SEISMIC DATA ACQUISITION AND PROCESSING

Kabir Roy Chowdhury
Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands

Definition

Seismic data acquisition

Generation of (artificial) seismic signals on land (on surface, or, buried) or in water, reception of the signals after they travel through the interior of the earth, and their (digital) recording for later analysis.

Seismic data processing

Analysis of recorded seismic signals to filter (reduce/eliminate) unwanted components (noise) and create an image of the subsurface to enable geological interpretation, and eventually to obtain an estimate of the distribution of material properties in the subsurface (inversion).

Introduction

Reflection seismics is akin to the "echo-in-the-well" experiment, it involves calculating the depth of the geological boundary from the two-way travel-time (TWT) of the seismic signal and its speed.

Seismic data acquisition and processing aims mainly to obtain an image of the sedimentary basins in interior of the earth, using waves generated by "artificial" earthquakes. These images can then be used to identify locations favorable for accumulation of hydrocarbons (oil and gas), which may then be drilled to determine the ground truth – and eventually to exploit the resources. Since the first known reflection seismic experiment in 1921 near Oklahoma City, USA (Figure 1.3, Sheriff and Geldart 1995), reflection seismics has established itself as the most accurate technique to image the sedimentary basins for the exploration of hydrocarbons.

The phrase "seismic" instead of "seismological" in the following stresses the "man-made" nature of the waves used. Both seismics and seismology use the basic theory

Seismic Data Acquisition and Processing, Table 1 Imaging the earth using natural/artificial earthquakes

Keyword	Seismics	Seismology
Wave source	Explosions, vibrations	Natural earthquakes
Energy penetration	Shallow	Deep
Max imaging depth	Base of crust	Whole earth
Location of source	Precisely known	Estimated post facto
Time of occurrence	Precisely known	Estimated post facto
Energy involved	Small-medium	Can be huge
Wave-propagation	Mostly vertical	Mostly horizontal
Frequencies mostly excited/used	1–100 Hz	0.01–1 Hz
Receivers	Geophones	Seismometers
Wave-field sampling	Dense	Sparse (getting better)
Data volume	Terabytes	Gigabytes
Accuracy	Large	Small–medium
Main application	Oil and gas	Earth-structure
Other applications	Civil engg., crustal	Civil engg.
Investment	$$$$	$$

of wave-propagation through the earth, for which Aki and Richards (2002) is a good resource. Table 1 summarizes the important differences between the two approaches though; let us briefly look at two.

Frequency vs. period: Due to the spectral range of the signals involved, seismology traditionally uses period (s) to describe the waves, whereas in seismics, frequency (Hz) is used. Waves provide information about the medium through which they propagate at the scale of their wavelength, use of higher frequencies in seismics (shorter wavelengths) leads therefore to a greater resolution (of the structure) compared to seismology.

Wave-propagation: seismology – again historically – mostly uses refracted energy, whereas exploration seismics is often synonymous with reflection seismics, although refraction seismic exploration predates the latter.

This essay will be mainly concerned with acquisition and processing of reflection seismic data. Note, however, that seismics is being increasingly applied to both shallower depths (high-resolution seismics) and crustal-scale studies down to Moho and beyond (deep seismics), see *Deep Seismic Reflection and Refraction Profiling* for details of the latter. Seismic data acquisition and processing is a broad subject, the treatment here will have to make choices based upon space constraints, etc. Some subtopics, e.g., "*Seismic, Migration*" are, however, dealt with in separate essays.

The reader is assumed to be familiar with the basic theory of elasticity and wave-propagation, and the related concepts of reflection, refraction, and scattering. The concept of rays will be frequently used – especially in illustrations – for convenience; real seismic signals are of course associated with wave-fronts. Similarly, the figures will depict a 2-D (section) of the 3-D earth.

There are many good resources available even for the narrower field of Reflection Seismic Data Acquisition and (Signal) Processing, e.g., Vermeer (2002), Yilmaz (2001), Menke (1989), Liner (2004), and Sheriff and Geldart (1995); the last one also contains some historical background and material over refraction seismics. Recently, some resources have also been made available for downloading on the internet, e.g., Claerbout (1985a, b).

In this article, all-capitals will be used to denote acronyms for jargons, of which there are quite a few (e.g., TWT above); phrases within double quotes will refer to an article elsewhere in this volume.

Seismic data acquisition

Before seismic signals could be processed, an artificial wave-field has to be generated using suitable *sources* at appropriate locations, measured by *receivers* at other locations after getting reflected back from within the earth, and stored using *recorders*. Design of a seismic survey (*geometry*) needs inputs from regional geology, exploration objectives, and logistical considerations.

At first confined to land, seismic surveys are now-a-days carried out mostly in marine environments in round-the-clock operations using large vessels and a lot of instrumentation; single-channel seismics has faded away in favor of multi-channel acquisition, allowing much more information to be obtained (see *Single and Multichannel Seismics*). Table 2 gives an overview of the equipments used under different field environments.

Seismic sources

One needs a signal that is high in energy (amplitude) to ensure a good depth penetration, and short in duration to ensure accurate determination and *differentiation* of the travel-times – a Dirac-Delta spike would be ideal, which, however, is a-causal. In practice, a sharp, compact, and *repeatable* signal is preferred. This quasi-idealized wave-form, possessing finite temporal duration and frequency band-width (both with respect to the ambient noise), is called a *wavelet*. The source wavelet changes form as it travels through the earth due to several physical processes to be briefly discussed below.

Repeatability of the source wavelet – that is, that of its *amplitude* and *phase* content – is an important prerequisite for the later processing steps. Explosives were the initial choice for source on land, providing large energy (good depth of penetration) but having non-repeatable signal shape and negative environmental impact. Development of large, truck-mounted electromechanical vibrators has led since 1960s to their increasing use in land-seismics, with both above disadvantages of impulsive sources reduced significantly.

In marine environment, compressed air is used – with explosion (*air gun*) or implosion (*water gun*) – to create the acoustic waves. The sources are towed by a ship together with the receivers (*single vessel seismic*), or, by a separate ship (*two-ship seismics*).

Seismic Data Acquisition and Processing, Table 2 Sources and receivers used in seismic surveys

Environment	Sources	Receivers
Land	Explosives/ vibrators/impact	Geophones
Marine	Air/water -guns	Hydrophones
Water bottom	Explosives/guns	Geo/hydro-phones
Onshore-offshore	Explosives/guns	Geo/hydro-phones

There have also been experiments with shear-wave sources – both impact-type and vibratory. These – either alone, or together with compressive sources – can provide extra information about the subsurface medium. For investigating shallower structures in engineering, environmental and archeological applications, small impact-based sources, e.g., weight-drops, rifles, etc., and even portable vibrators are being frequently used, and provide the required higher resolution.

Seismic receivers

All land-seismic (and seismological) receivers (*geophones, seismometers*) are electromechanical devices, that transform the relative motion of the medium in which they are embedded, into electrical voltages. Fidelity of this transformation, both in *amplitude* and *phase*, is important to ensure maximum information retention for later retrieval – a flat amplitude response, with no phase distortion within the band of frequencies that are of interest, would be ideal. The output of a geophone can be made to be proportional to the displacement/velocity/acceleration associated with the causative motion. Dennison (1953) provides an overview of the physico-mathematical underpinnings of geophone design; see also *Seismic Instrumentation*.

Originally, geophones were designed to move – and hence record information – only in the vertical direction. Later, the importance of recording and analyzing the entire three-dimensional elastic wave-field came to be realized. Multi-component receivers, enabling recording/analysis of both horizontal components, or, all three spatial components of the ground movement are being increasingly used even in large-scale surveys.

For use in water, small piezoelectric elements – hydrophones – are employed to record pressure variations – modern deployments typically consist of thousands of such elements being towed near the water surface by *streamers*, several kilometers long, which are liquid-filled plastic tubes, fitted with fins (for buoyancy), gps receivers (for location information) and fiber-optic cables (to transfer the data) to the ship.

Finally, three-component receivers may be deployed together with hydrophones at the water bottom (4C), to record the wave-field across it, see *Ocean Bottom Seismics*.

Note, that both sources and receivers may be deployed in *groups*, using specific patterns, which affect the generation and sampling of the wave-field due to their direction-dependent radiation/reception characteristics.

Seismic recorder

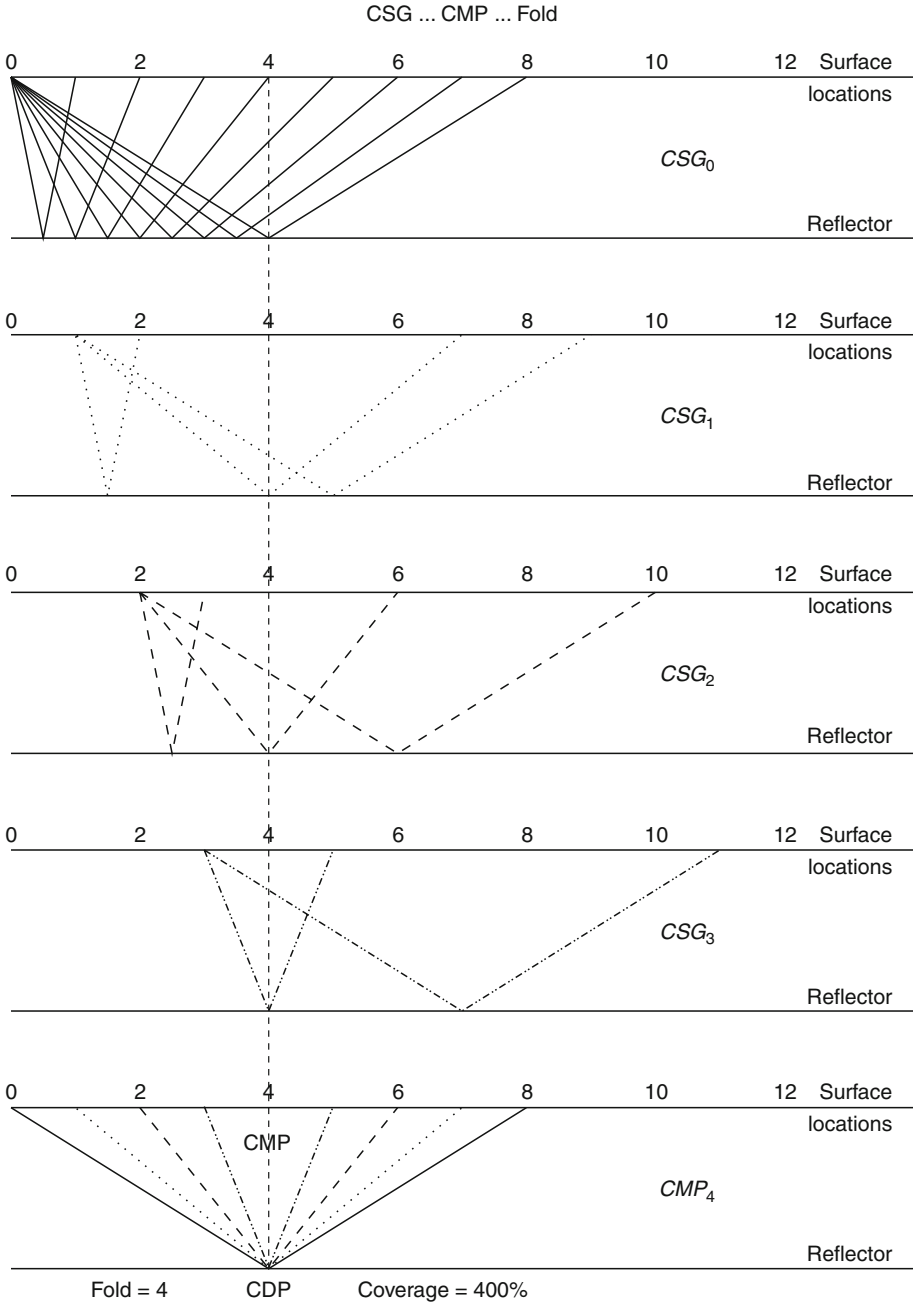
The time-varying electrical signals output by the receivers represent *arrivals* back-scattered from different depths, all juxtaposed in time and embedded within the ever present background noise, and require storage for later processing. In the beginning, photographic films and later magnetic tapes were used for this purpose. The digital revolution starting in the 1960s, itself partly driven by the needs of seismic data acquisition and processing, caused a complete shift to in-situ digitization and digital storage. Similarly, the wires connecting the receivers to the recorder have been mostly replaced by fiber-optic cables or wireless. Preserving the frequency, amplitude, and phase of the signal and the desired *dynamic range* are important considerations in designing the digitizing (*sampling*) unit. As each *receiver (group)* corresponds to a different *channel* in the recorder, digitization in such systems (typically consisting of thousands of channels) must preserve the time-base, to enable comparison of the arrival times between different *traces*. Also, the actual time of the *shot*, t_0, must be transferred from the source to the recorder and recorded; it is often used to start the recording process itself, as seismics is only interested in travel-times, i.e., arrival-times with respect to t_0. The digitized data – uniformly sampled time-series – from each individual experiment (*shot*), consisting of multiple *traces* (output of receivers), is called a *seismic record*.

Acquisition geometry

Assuming a *layer-cake* model (sedimentary beds parallel to the surface), early surveys deployed a number of sources and receivers along a straight line on the earth-surface to obtain a vertical cross-section of the geology below this line (2-D).

Figure 1 shows schematically the approach in such a survey, and is – in spite of its simplifications – useful in understanding several basic ideas. All the five panels show the earth-surface at the top, and a reflecting boundary (*target*), parallel to it, at some depth. Numerals on the surface represent surveyed equi-distant flag-positions to denote locations. The top panel shows the first measurement, with the source at "0" and eight receivers at locations "1" through "8". Assuming a homogeneous and isotropic medium, the paths predicted by Snell's law for a part of the source energy to first travel downward to the target and then reflect upward from it to reach the receivers are indicated by the oblique lines.

The signals output from the receivers are digitized in the recorder to yield a *seismic record*, i.e., a collection of *seismic traces*. Such an ordered collection of seismic traces is called a *gather*. Having a common *source*, the record resulting from our first measurement is a *common source gather* (CSG_0), the suffix denoting source position.



Seismic Data Acquisition and Processing, Figure 1 Schematics of seismic data acquisition by common-mid-point (CMP) profiling. Panels CSG_0 through CSG_3 represent common source gathers (CSG); CMP_4 is a common mid point gather for one common depth point (CDP). See text for details.

Under the twin-idealizations of no background noise, and a spike-like source signal, each channel in the recorder (*seismic trace*) will consist of one single blip corresponding to the arrival time of the signal; in reality, the arrivals will have random background oscillations due to noise, and one *wavelet* corresponding to the single reflection arrival. Assuming constant speed of propagation v and depth to the target H, it is trivial to show (e.g., Sheriff and Geldart, 1995) that the travel-times to the receivers can be written as $t_x^2 = (x^2 + 4H^2)/v^2 = t_0^2 + x^2/v^2$, t_x being the arrival time recorded by a receiver at a source-receiver *offset* of x. The travel-time curve for such a situation is thus a hyperbola – this simple relationship underlies much of seismic processing. t^2 plotted against x^2 thus yields a straight line, the slope being v^{-2}, i.e., square of the

slowness of the medium. Note that in seismics, the *velocity*, which sensu stricto is a vector, is almost always used to denote the local wave *speed* (a scalar), which is a property of the medium (rocks) . . . we shall follow this usage.

Our aim is to find H, the depth to the target ($= t_0/2v$). We have thus to estimate t_0 from the rest of the reflection hyperbola. Note that the reflection points on the target for the different receivers are different.

Hence, in what has become almost universal practice, the measurement is repeated after shifting the whole set-up laterally along the measurement line, keeping all the relative distances the same. In panel 2 of Figure 1, the source and the receivers have been shifted right by one unit; only a few ray-paths are shown for this gather (CSG_1). Similarly, gathers CSG_2 and CSG_3 are also measured and recorded. During these measurements, the same receiver locations recorded signals from different sources so that a post-measurement re-arrangement of the traces could also yield *common receiver gathers* (CRG); in our case, we would obtain $CRG_1 - CRG_{11}$. These are useful for certain processing situations.

The lowest panel of this figure shows a special kind of re-sorting, collecting the traces from the four *shots* with one *common reflection point* (CRP). Four traces corresponding to source-receiver combinations of 0/8, 1/7, 2/6 and 3/5 were selected respectively from the four gathers. For our simple geometry, the four ray-paths shown share two things – a *common mid point* (CMP_4) between their respective source and receiver locations and the *common depth point* (CDP) at the target depth, the latter being the same as CRP. Such a gather is called a CMP-gather, and indexed by the position of the CMP. The travel-time plot of the reflection arrivals in a CMP-gather is also a hyperbola.

The four ray-paths shown for the gather CMP_4 all have the same reflection point, and thus contain information about the same subsurface geology. The arrival times of the reflection signal in the four traces are of course different, as the travel paths are different. If this difference is corrected for, then adding the four traces should increase the coherent signal (information regarding the CRP) with respect to the *random* noise. The improvement of S/N by adding N traces is given by

$$\frac{\sum(\text{N traces with identical signal})}{\sum(\text{N traces with random signal})} \approx \frac{N}{\sqrt{N}} = \sqrt{N}.$$

The improvement of the signal-to-noise (S/N) ratio is thus roughly proportional to the square-root of the number of traces added. This number (4 in our case) depends upon the survey geometry, and is called the *fold* of the survey. Starting from fold 1 for $CMP_{0.5}$ (not shown), it gradually builds up to its nominal value (4 in this case), and again drops-off at the other end of the survey.

Acquisition configuration can be specified by expressions describing the position of the source relative to the receivers, viz., *end-on, split-spread, broad side*, etc. Depending upon the geology and the noise regime, these configurations, as also varying *fold*, leave subtle but important *footprints* on the data.

In reality, the geology is of course not as in Figure 1, presence of structure (*dips, faults, folds*, etc.) is what makes hydrocarbon accumulation possible in the first place. Processing of 2-D data can remedy this situation – though only partially. Availability of more equipment and data processing power led therefore to development of 3-D acquisition, with receivers laid out on the surface in a 2-D pattern, and sources also positioned in a different 2-D pattern, thus causing a better illumination of the subsurface by the seismic waves. Here too the basic concept of adding *fold* number of traces in a CMP-gather holds sway – point-shaped CMPs and CDPs being replaced by finite *bins*, their sizes depending upon the survey design and objectives (see Vermeer, 2002 for further insight into acquisition design).

In areas with structural complexity, the simplifying assumptions of CMP-processing break down, and the availability of computer power may make it possible – nay desirable – to process each trace of the recorded CSG separately, to try to obtain a better image (see also *Seismic Imaging, Overview* and *Seismic, Migration*).

Restricting the deployment to the surface of the earth implies – as we shall see later – a bias for horizontal structures; this was eventually removed by carrying out measurements inside bore-holes called VSP; see *Vertical Seismic Profiling* for details. Finally, better recording instrumentation coupled with the need to detect changes in the hydrocarbon reservoirs resulting from exploitation has given rise to *time lapse seismic* (4D), whereby repeat imaging of the same area, carried out after several years of production, is used to validate/improve production models for reservoirs.

Seismic data processing

Introduction

Reflection seismic data, acquired in the field, has to be taken through several processing steps, before it can be interpreted in terms of the subsurface structure. The source signal, on its way down, and back up to the receivers is modified by many factors; the aim of processing is to *undo* (i.e., correct for) as many/much of these effects as possible, leaving only the effects due to the causative structure of interest (geology) to be interpreted.

Seismic data is a spatio-temporal sampling of the back-scattered seismic wave-field, an ordered collection of traces, and can be considered to be a 2-D or 3-D data matrix along with some auxiliary information regarding location, etc. The traces themselves are an ordered collection of uniformly sampled amplitude values (*time-series*), with relevant information contained in their respective headers in (internationally) agreed formats. All processing steps aim to improve the spatio-temporal S/N ratio of the data by reducing the noise and/or by sharpening the wave-form (to improve the resolution).

Signal versus noise

Before proceeding further, it is useful to reflect on the terms *signal and noise*. That it is a matter of perspective is clear from this relative definition: *signal is useful noise and noise is useless signal*. In other words, someone's noise is someone else's signal, and vice-versa. For example, the *ground-roll*, hated in reflection seismics, is useful in surface-wave seismology and shallow-seismics. Amazingly, using noise for seismic imaging has now become a field of active research (see the section "*Seismic Noise*" for references).

In reflection seismics, signal is synonymous with *primary reflection. Primaries*, as these are often referred to, represent seismic energy reflected only once during its travel from source to receiver. Everything else, present in the traces, is taken to be *noise*. This includes multiply reflected energy (*multiples*), diffractions (caused by *sharp* structures in the subsurface, e.g., faults, pinch-outs), refracted arrivals, surface waves (*ground-roll*). Non-geological noise sources include nature (wind, waves, animals) and man (traffic, industry, etc.). From processing point of view, noise could be coherent (ground-roll, water-pump, multiples), or, incoherent, each needing a different strategy. See *Seismic Noise* for details.

Kinematics of the seismic signal (primaries)

Starting with some simple (but inaccurate) assumptions, e.g., horizontal layering, constant speed, etc., useful structural information can be extracted – a large data volume contributing to the robustness of the processing algorithms (also see *Seismic, Migration*). In this section, we focus on the travel-times of the waves (visualized as *rays*), see also additional information in "*Seismic, Ray Theory*."

NMO

The travel-time for a primary reflection from a horizontal reflector, shown earlier to be hyperbolic, can be rewritten as:

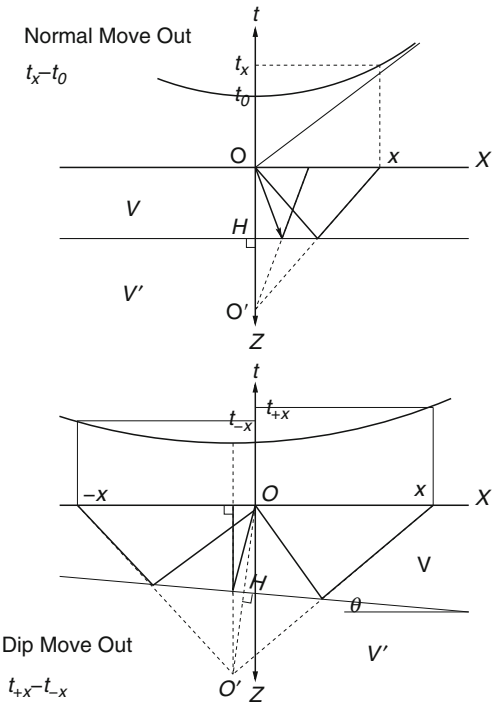
$$t_x - t_0 = \frac{\sqrt{x^2 + 4H^2} - 2H}{v}. \tag{1}$$

The quantity on the left is the difference (see Figure 2) between the oblique reflection-time at source-receiver offset (distance) x and the vertical TWT, and leads to the relation:

$$\Delta t_x = \frac{2H}{v}\left(\sqrt{1 + \frac{x^2}{4H^2}} - 1\right) = t_0\left(\sqrt{1 + \frac{x^2}{4H^2}} - 1\right). \tag{2}$$

Expanding the expression under square-root, and recognizing that in most seismic measurements offset $\ll$ target depth, we obtain the approximate relation (Equation 3), which could be improved by retaining additional higher order terms.

$$\Delta t_x \approx \frac{x^2}{4vH} = \frac{x^2}{2v^2 t_0} \tag{3}$$



Seismic Data Acquisition and Processing, Figure 2 NMO of a primary reflector; horizontal (above), dipping (below). See text for details.

Δt_x ($= t_x - t_0$) is called the *normal move-out* (NMO) associated with the reflection travel-time. NMO can be used to align the primary reflection in all traces at t_0 (TWT) by removing the effect of source-receiver distance (offset), i.e., by *flattening the reflector*. NMO, an important concept in seismics, is used both to first identify *primaries*, and later to align them for imaging the reflector. Note that to use 3, we need to know x (source-receiver offset), v (speed), and H (target depth); in practice, x is known and iteration is used to obtain optimal values for v and H.

Dipping bed

For a dipping reflector (Figure 2), travel-time for the primary reflection is still hyperbolic, given by

$$v^2 t_\theta^2 = x^2 + 4H^2 + 4Hx\,\sin\theta. \tag{4}$$

The minimum of the hyperbola is now shifted *updip*; the quantity $t_{+x} - t_{-x}$ is a measure of the asymmetry, and can be used to estimate the dip.

Many reflectors: layer-cake

Dix (1955) considered the case of many reflectors parallel to the surface – a good starting model for sedimentary sequences – and showed, that here too, the travel-time curve can be approximated at short offsets by a hyperbola:

Seismic Data Acquisition and Processing, Table 3 Jargons associated with the term velocity in seismics

Jargon	Brief description
v_{int}	Speed in a geological interval (assumed constant)
v_{av}	Average speed between two points along a ray path
v_{app}	Apparent speed measured by receivers in field (= *dx/dt*)
v_{nmo}	Speed used for NMO correction (strictly, only for one layer)
v_{rms}	Dix' root-mean-square NMO velocity for layer-cake situation
v_{stk}	*Best* velocity to stack CMP-gathers
v_{mig}	*Best* velocity to migrate the seismic data

$$t_x^2 \approx t_0^2 + \frac{x^2}{v_{\text{rms}}^2}, \quad \text{with} \quad v_{\text{rms}} = \sqrt{\frac{\sum v_i^2 \Delta t_i}{\sum \Delta t_i}}. \tag{5}$$

The homogeneous velocity v (= v_{nmo}) is now replaced by v_{rms} (root-mean-square velocity), which depends upon the velocities of the layers v_i and the vertical transit times t_i through them. v_{rms} plays a role similar to v_{nmo} in flattening the primaries in the multi-layer case. Individual layer-velocities may then be computed from the Dix' equation:

$$v_n = \sqrt{\frac{v_{\text{rms},n}^2 t_n - v_{\text{rms},n-1}^2 t_{n-1}}{t_n - t_{n-1}}}. \tag{6}$$

Velocities in seismics

In seismics, different terms are used to denote "velocity" depending upon the context. Table 3 lists a few, along with brief explanations, some of these will be elaborated later.

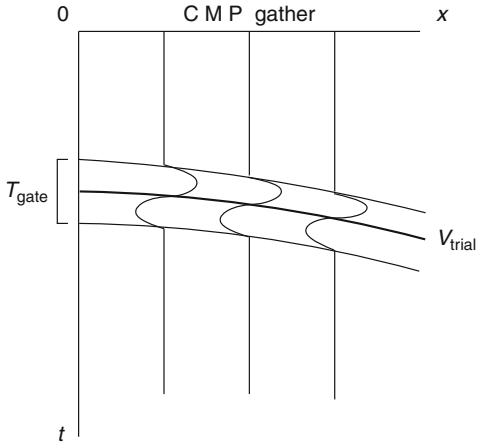
NMO stretch

After NMO correction, a time-interval Δt_x, say corresponding to a period of the wavelet recorded on a trace at an offset x, becomes Δt_0; the wavelet is thus distorted (*stretched*). The expression 100 $(\Delta t_0 - \Delta t_x)/\Delta t_x$ is a percentage measure of this stretch – 0% implying no distortion. In practice, a threshold percentage is specified to exclude parts of data – relatively largeoffsets and small arrival times – from being NMO corrected (and taking part in further processing).

Semblance: a measure of signal alignment

To apply optimal NMO correction, a quantitative measure of alignment of amplitudes, across several traces, is useful. Such a measure of similarity between n (amplitude) values, called *semblance*, is defined by

$$S = \frac{\left(\sum_n val\right)^2}{n \sum_n val^2}, \quad \text{and}$$
$$S_{\text{gate}} = \frac{\sum_{\text{gate}} \left(\sum_n val\right)^2}{\sum_{\text{gate}} \left(\sum_n val^2\right)}. \tag{7}$$



Seismic Data Acquisition and Processing, Figure 3 Schematical drawing showing calculation of multi-channel semblance. *Curved bold line* represents the move-out curve for a trial velocity, the *two surrounding lines* represent the boundaries of the time-gate; see text for details.

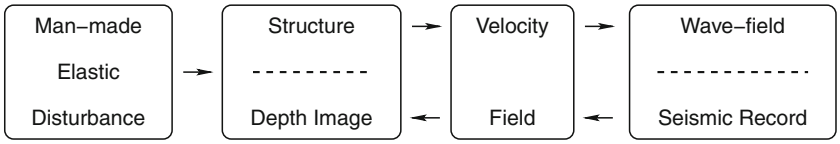
Note that semblance is a dimensionless number between 1 (perfect match), and is 0 (perfect mismatch). The second form uses a time-gate along the traces, generally having the width of the dominant period of the signal, for increased robustness. Semblance is used extensively in modern *reflection velocity analysis*, to evaluate the goodness of alignments of primary reflections along move-out curves computed for a range of trial velocities (Figure 3).

Velocity: processing point of view

Wave-speed (called velocity in seismics) in the medium is the missing link needed to convert the travel-time information to depths required for structural interpretation – and eventual drilling. Note that *velocity* is needed to find the structure (geology), but structure is needed to find the velocity. This catch-22 situation is solved iteratively – shown schematically in Figure 4.

Velocity is a macroscopic (wavelength-scale average) property of the rock depending upon the density and elastic properties of the minerals making up the lithology (see *Seismic Properties of Rocks*). In rocks of interest in seismics (*sandstone, shale, limestone*), velocity is not a good indicator of lithology, with considerable overlap in values, with some exceptions, e.g., salt, anhydrite (relatively higher velocity). Presence of *porosity and pore-fluids (water, oil, gas)* is the most important factor for this overlap, and is in turn caused by the *burial history* of the rocks. Wave-propagation in fluid-filled porous media is described by Biot-Gassman theory, see Lee (2008) for references and recent developments.

Propagation velocity (the missing link) can be estimated by direct measurements (see, e.g., Sheriff and Geldart, 1995 for details), which have shortcomings though (see Table 4). The velocity used for processing



Seismic Data Acquisition and Processing, Figure 4 Iteratively solving for both structure and velocity in seismics.

Seismic Data Acquisition and Processing, Table 4 Direct determination of seismic velocities and their shortcomings

Method	Shortcoming
Uphole-time	Useful only for the weathering layer
Check-shots, well-shoot	Limited depth-range, destructive
VSP	Available late in exploration, expensive
Sonic log	Available late, noisy (high-frequency)
Lab measurements	Limited availability

seismic reflection data is usually determined iteratively from the data itself, and will be described later.

Amplitude changes along the propagation path

Several factors cause the amplitude of the seismic waves to change as they travel from source to receiver. These can be corrected for, so as not to mask the weaker changes (signals) of interest.

Geometrical spreading

Conservation of energy requires a continuous reduction of amplitude, as a seismic wave-front spreads through a medium – hence the term *geometrical spreading*. The loss depends upon the mode of spreading and the distance traveled (r). For primaries (body-waves), amplitude ($\propto \sqrt{\text{energy}}$) decreases $\propto r^{-1}$, whereas for ground-roll (surface-wave), the decrease is $\propto r^{-1/2}$, the latter shows why ground-rolls, with their (relatively) large amplitudes, are a big problem in seismics.

Absorption

The propagating wave continuously loses energy due to absorption too, which is a physical property of the medium, and can be described by several equivalent parameters, e.g., *absorption coefficient, damping factor*, etc., the most common being the *quality factor* of the material $Q = 2\pi/$[fractional energy lost per cycle]. It is a dimensionless quantity, with a value 0 implying perfect absorption and ∞ implying perfect elasticity. Absorption, with Q considered to be frequency-independent within the band-width of interest in seismics, causes relatively greater attenuation of higher frequencies – leading to a change in the wave-form during propagation. See *Seismic, Viscoelastic Attenuation* for more details.

Energy partitioning at interfaces

Boundaries of geological heterogeneities (layering, faults, etc.) also cause changes in the amplitude of the wavelet; such changes are, indeed, of prime interest in seismics. As in optics, the interaction between the wave-fronts and the geological structure depends upon their relative dimensions, i.e., their radii of curvature-with specular reflections and point-scattering building the two end-members, both of which are encountered in seismics. Another concept from optics, *diffraction*, is useful to understand the complexity of the interaction between the wave-front and the medium. See *Seismic Diffraction*; *Seismic Waves, Scattering*, and *Energy Partitioning of Seismic Waves* for additional details.

Waveforms: convolution, deconvolution

Factors modifying the source signal along the path of the seismic wave may be divided as: near-source (*ns*), i.e., weathering layer, earth (*e*), i.e., the target geology, near-receiver (*nr*), receiver (*r*), and recorder (*rec*), with the output trace (o) as the final result. Each of these, denoted in Equation 8 below by the expression in parentheses, affects (*filters*) the source wavelet (s). In a series of papers/reports (Robinson, 2005; Treitel, 2005), the MIT geophysical analysis group (GAG) laid the foundation of the *digital revolution* in seismic data processing, by examining the nature of these filters and developing methods to undo their effects. These resulted in major advances in *time-series analysis and digital filtering* (Robinson and Treitel, 1964), and a critical evaluation of the (statistical) nature of earth's reflectivity (*target geology*).

Convolutional model of the seismic trace

As the source- and recorded-signals are both time-series (uniformly sampled, ordered collection of amplitudes), it is useful to represent all the other elements mentioned in the above-paragraph also as such. For a column of vertically layered reflectivity, such a time-series would correspond to values equal to *RC*s placed at times converted from depths using velocities. Now, making the crucial assumption that all these filter elements are linear systems, the recorded trace can be expressed as:

$$o(t) = s(t) * ns(t) * e(t) * nr(t) * r(t) * rec(t) + n(t). \tag{8}$$

In Equation 8, $*$ (*star*) is the convolution operator, well-known in the theory of linear systems; $n(t)$ represents some additive noise which does not follow this model, hopefully, it is mostly removed early in the processing.

The time-series that transform $s(t)$ into $o(t)$ can also be interpreted as the *impulse response* of the corresponding elements, e.g., $r(t)$ is the response of the receiver to a sudden spike signal. Using *Fourier Transforms* to change the time-series into their *spectra*, and remembering that convolution in time-domain corresponds to multiplication in frequency domain, one obtains:

$$O(\omega) = S(\omega) \cdot NS(\omega) \cdot E(\omega) \cdot NR(\omega) \cdot R(\omega) \cdot REC(\omega), \tag{9}$$

where the noise term has been neglected (see Sheriff and Geldart, 1995 for introduction to linear operators and Fourier theory). Equation 9 clearly shows the *filtering* effect of the different elements, each one modifying the spectrum of the incoming signal by modifying/removing a part of its frequencies. Our aim, in seismic data processing, is to extract $e(t)$, the geological structure, from the recorded signal $o(t)$.

Deconvolution as inverse filtering

Undoing the act of the filterings implied in Equations 8 and 9 is called *deconvolution* (decon), or, *inverse filtering*. Equation 9 can be rewritten as $O(\omega) = \mathrm{E}(\omega).\,\mathrm{REST}(\omega)$, where $\mathrm{REST}(\omega)$ groups together all the elements on the right besides the geology. Then, $E(\omega)$, or $e(t)$, can be *estimated* from

$$E(\omega) \cong O(\omega)/REST(\omega), \quad \text{or,} \qquad e(t) \cong o(t) * rest(t)^{-1}. \tag{10}$$

The approximation sign, for both forms of Equation 10 – in frequency domain (first), or, in time-domain (second) – is necessary, even in the *noise-free* case. Spectral division needs precautions to avoid zero-division in parts of the spectrum, where frequencies have been weakened/removed. Fortunately, addition of *noise* helps, since signals of interest in seismics exist – by definition – only above the ambient noise level. See Liner (2004), Sheriff and Geldart (1995) and Yilmaz (2001) for the stabilizing role of *spectral whitening* in decon.

Wavelet processing

Wavelets: Let's take a closer look at *seismic wavelet*, introduced in the section about seismic sources, as a signal of finite frequency band-width and temporal duration. Using standard concepts from time-series analysis (Sheriff and Geldart, 1995; Yilmaz, 2001), simple examples of wavelets are:

$$a : (3, -2, 1), \quad b : (2, 3, -1) \quad \text{and} \quad c : (-1, 2, 3),$$

the numbers representing uniformly sampled amplitudes starting from $t = 0$. Remembering that squares of the amplitudes in a wave(let) are measures of energy, we see that these three wavelets, while looking very different, have the same total energy. Depending upon the energy build-up, wavelet a is called *minimum delay* (energy is front loaded), b is *mixed delay*, and c is *maximum delay*; physical (causal) wavelets are minimum delay, although in the example, a is not strictly causal, due to the instantaneous build-up of energy at $t = 0$. In frequency domain, the expressions minimum/mixed/maximum-phase are used instead.

Wavelet estimation: *Auto-correlation* of the wavelets a-c are all symmetrical about $t = 0$, i.e., have no phase information, e.g., $\phi_{bb} = (-2, 3, 14, 3, -2)$; these are Fourier Transforms of the respective power-spectra. In seismics, an estimate of the power spectrum is often available from the data. The question then arises whether an estimate of the wavelet may be obtained from it – an outline follows. Using Z-transform notation, one can write the wavelet, say c, and its auto-correlation as polynomials:

$$C(Z) = -1 + 2Z + 3Z^2, \text{ and,}$$
$$\Phi_{cc}(Z) = -3Z^{-2} + 4Z^{-1} + 14 + 4Z - 3Z^2,$$

Z being the *unit-delay* operator, its powers denoting time-shifts with respect to $t = 0$. According to the fundamental theorem of algebra, a polynomial of degree n in Z must have n roots, i.e., it can be expressed as a product of n factors of the form: $(Z - Z_1)(Z - Z_2) \ldots (Z - Z_n)$, each factor representing a basic wavelet (*doublet*). Half the doublets of an auto-correlation polynomial are minimum delay, their product represents the Z-transform of the unique minimum delay causative wavelet. See Yilmaz (2001) and Sheriff and Geldart (1995) for details, assumptions, critical remarks, and alternate approaches (e.g., *homomorphic deconvolution*) to deconvolution of time-series.

Wavelet manipulation: Much of seismic processing is involved with manipulating the wavelet (deconvolution in a general sense). While very powerful, it contains potential for pitfalls, if applied without a proper understanding of the suitability of the particular technique, as each decon step also causes artifacts.

Spiking decon aims to sharpen the shape of the signal, to improve temporal resolution – and interpretation. Ideally, it involves convolving the wavelet with its inverse operator, to yield a spike, i.e., perfect resolution.

Zero-phasing converts the signal to one with zero-phase; the result is a symmetrical signal (a-causal), and is primarily useful for interpretation if the peak can be made to coincide with the reflecting boundary.

Any-phasing is used in merging seismic datasets of different vintages and with differing source wavelets.

General shaping groups methods to convert the signal to any desired shape optimally – using some statistical criteria.

Depending upon whether a model is available for decon, the methods could also be divided in *deterministic*, i.e., model-based and *statistical*.

Deterministic deconvolution

Vibroseis processing. Vibrators (see the section on sources) use a repeatable source signal, called *sweep*. It is a time-limited (typically, 10–20 s long) signal with the frequency continuously varying between given start- and end-values, and comes in many flavors, e.g., up-, down-, linear-, non-linear-sweeps. Neglecting other terms, one could write from Equations 8 and 9: $o(t) = s(t)^* e(t)$. The recorded signal is thus the convolution of *earth reflectivity* with the sweep signal. We could remove its effect (deconvolve) by *cross-correlating* the observed signal with the sweep (which we know precisely), a process, which is equivalent to convolving with its time-reversed version, and get

$$s(-t) * o(t) = s(-t) * s(t) * e(t) \approx \delta(t) * e(t) \qquad (11)$$

Due to the sweep signal being time-limited, its auto-correlation is not a Delta-spike (ideal), but is a symmetrical (zero-phase) signal called *Klauder wavelet*. The result is thus not quite the desired earth reflectivity (although it has the correct phase) and needs further processing for improvement (see Yilmaz, 2001; Liner, 2004; Sheriff and Geldart, 1995).

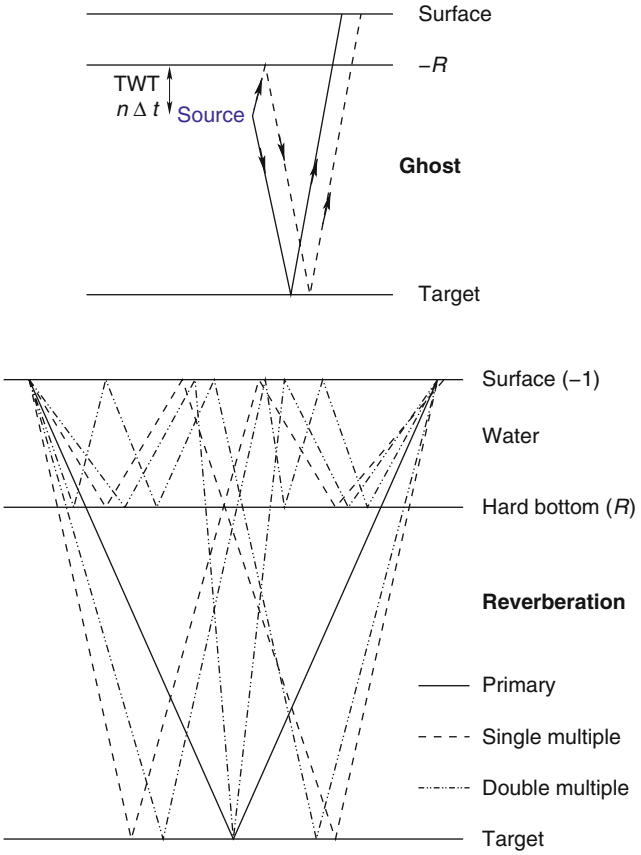
De-ghosting. The effect of large *RC*s in the shallow sub-surface has been mentioned earlier. Figure 5 shows one such situation; here the source is placed below the weathering layer, for better energy transmission toward the deeper target (ray going directly downward). A part of the wave-energy, also travels upward, and gets reflected down from the base of the weathering layer (ray going first up, and then down). In certain cases, the $RC_{\text{weathering}}$ could be quite large and negative. The energy reflected downward follows with a short delay behind the direct wave, and is called a *ghost*; the observed record is thus corrupted by that caused by a delayed ghost. Removing the latter from the recorded trace is called *deghosting,* and is an example of model-based decon. Assuming the TWT between the source and the base of the weathering to be n samples ($= n\Delta t$), one can write:

$$o(t) = s(t) - Rs(t - n\Delta t), \text{ or, using Z} - \text{transforms,}$$
$$O(Z) = S(Z) - RS(Z)Z^n = S(Z)(1 - RZ^n).$$

$(1 - RZ^n)$ is, clearly, the Z-transform of the ghost-operator. Hence, $S(Z) = R(Z)\ (1 - RZ^n)^{-1}$, or, $s(t) = o(t) + s(t - n)$. The last form above implies recursive filtering in the time-domain to achieve deghosting. Alternately, expanding $(1 - RZ^n)^{-1}$, the inverse-filter operator in time-domain can be written as

$$g(t)^{-1} = (1, 0, 0, \ldots + R, 0, 0, \ldots + R^2, 0, 0, \ldots)$$

De-reverberation. The lower part of Figure 5 shows another situation, where strong reflectivity associated with the water bottom causes long trains of high-amplitude *reverberation* of signals in the water layer. The ray-paths



Seismic Data Acquisition and Processing,
Figure 5 Deterministic deconvolution applied to ghost (above) and reverberation (below). The near vertical ray-paths are shown obliquely for better visualization, see text for details.

shown schematically are: one primary reflection from the target, two multiples reflected once in the water layer, and three multiples reflected twice; there could be many more, posing a serious problem in marine seismics. Depending upon the depth of water, certain frequencies will, as a result, experience severe distortion (enhancement or suppression). In the simplified case of a water-column with a TWT equal to the sampling interval, and remembering that the negative reflectivity causes phase-change, the total operator (signal + reverberation) can be written as:

$$w(t) = (1, -2R, +3R \ldots) \rightarrow W(Z)$$
$$= 1 - 2RZ + 3R^2Z^2 - \ldots = (1 + RZ)^{-2}.$$

It follows that the deconvolution in this case can be achieved by the operator

$$W(Z)^{-1} = (1 + RZ)^2, \quad \text{or,} \quad w(t)^{-1} = (1, 2R, R^2).$$

This elegant operator is called the Backus filter (see Backus, 1959).

Statistical deconvolution

In the absence of a deterministic model, one could attempt to change the signal wavelet to any desired shape, by designing filters that are optimal in a statistical sense. Based upon work in information theory by Norbert Wiener and others, the applications in seismics were pioneered by the MIT-GAG group, e.g., Robinson (1967). Schematically, the basic approach is:

assume filter → compute output → compute error →
minimize to get normal equations → solve for filter coefficients

If errors are assumed to be Gaussian, and l_2 norms are used, the operators obtained are called Wiener filters. Such *optimum filters* are used widely, e.g., in:

- Zero-lag spiking – to increase resolution
- Zero-phasing – to ease interpretation
- Prediction filtering – to remove multiples which are predictable, the remnant being the prediction error, corresponding to the deeper signal

Wiener optimum filter. The normal equations for the filter coefficients f are given by the matrix equation shown in Equation 12 in its compact form

$$\phi_{\text{input,input}} * f = \phi_{\text{input,output}}, \tag{12}$$

which relates the auto-correlation of the recorded (input) wavelet to its cross-correlation with the desired (output) wavelet. For the derivation of Equation 12, and a detailed treatment of statistical deconvolution, see, e.g., Yilmaz (2001) or Sheriff and Geldart (1995) – an example is shown below to illustrate the approach.

Spiking filter. If the wavelets are all n-sample long, the Matrix Equation 12 can be expanded as

$$\begin{pmatrix} \phi_{i,i}(0) & \phi_{i,i}(1) & \dots & \phi_{i,i}(n-1) \\ \phi_{i,i}(1) & \phi_{i,i}(0) & \dots & \phi_{i,i}(n-2) \\ \phi_{i,i}(n-1) & \phi_{i,i}(n-2) & \dots & \phi_{i,i}(0) \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_{n-1} \end{pmatrix} = \begin{pmatrix} \phi_{i,d}(0) \\ \phi_{i,d}(1) \\ \phi_{i,d}(n-1) \end{pmatrix} \tag{13}$$

The auto-correlation matrix $\phi_{i,i}$ in Equation 12, with the same element in each diagonal descending from left to right, is a Toeplitz matrix; f is a column vector with the filter coefficients to be determined and $\phi_{i,o}$ is a column vector with elements from the cross-correlation matrix. Equations with Toeplitz matrices can be efficiently solved by a procedure called *Levinson recursion*.

Wiener filter: a simple example. Given the input wavelet $i_t = (1, -1/2)$, let us find the optimum two-element Wiener-operator to transform it to the desired wavelet $d_t = (1, 0)$, i.e., a zero-delay unit-spike. We get $\phi_{i,i} = (5/4, -1/2)$, and, $\phi_{i,d} = (1, 0)$. Equation 13 then becomes

$$\begin{pmatrix} 5/4 & -1/2 \\ -1/2 & 5/4 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{yielding} \quad f_{\text{Wiener}} = \left(\frac{20}{21}, \frac{20}{21}\right). \tag{14}$$

Applying this filter to the input, we obtain the output (20/21, −2/21, −4/21), which compared to the desired output, gives a squared error of 1/21. The ideal filter for this decon is the *inverse filter* for the input wavelet. Writing $I(Z) = 1 - Z/2$ for the Z-transform of the input, the Z-transform of the inverse filter (which will convert the input to an ideal unit-spike) is $= (1 - Z/2)^{-1} = 1 + Z/2 + Z^2/4 + \ldots$, which is an infinitely long operator! For an honest comparison of its performance with that of the Wiener filter, we apply its first two terms to the input, getting the filtered version as (1, 0, −1/4); although looking better at the first glance, its squared error is 1/16, i.e., larger than that of the Wiener filter! It can be shown that the Wiener filter is the *best* two-element filter for this problem.

Suppose the input wavelet is (−1/2, 1), i.e., not minimum delay, which we want to transform to a zero-delay spike. Normal equations now give the Wiener filter as (−10/21, −4/21), with the output (5/21, −8/21, −4/21) and the squared error as 6/21. Inverse filter is now (−2, −4, −8, ...), which is extremely unstable! Its first two filter elements give the output (1, 0, −4) with 16 as error! Wiener filter performs here worse than in the first case, because, it was trying to convert a maximum delay wavelet to a minimum-delay spike, but it still does better than the (finite) inverse filter. In this case, if a maximum delay spike (0, 1) was desired, Wiener filter coefficients would be (16/21, −2/21), giving a filtered output of (−8/21, 17/21, −2/21) with a squared error 4/21, which is better than that for a zero-lag spike output. Table 5 summarizes the results.

The processing flow: putting it all together

Most of the processing *modules* (*filters*) operate on the data (time-series) sequentially, the entire process resembling a *flow*, though there are a few stand-alone modules

Seismic Data Acquisition and Processing, Table 5 Performance of Wiener and inverse filters

Input wavelet	Desired wavelet	Two-point Wiener		Two-point inverse	
		Filter	Error	Filter	Error
(1, −0.5)	(1, 0)	(20/21, 8/21)	1/21	(1, 0.5)	1/16
(−1/2, 1)	(1, 0)	(−10/21, −4/21)	6/21	(−2, −4)	16
(−1/2, 1)	(0, 1)	(16/21, −2/21)	4/21		

too. The operations could be on individual traces (*single-channel*), or on a gather of traces (*multi-channel*). Schematically, a seismic processing flow looks like:

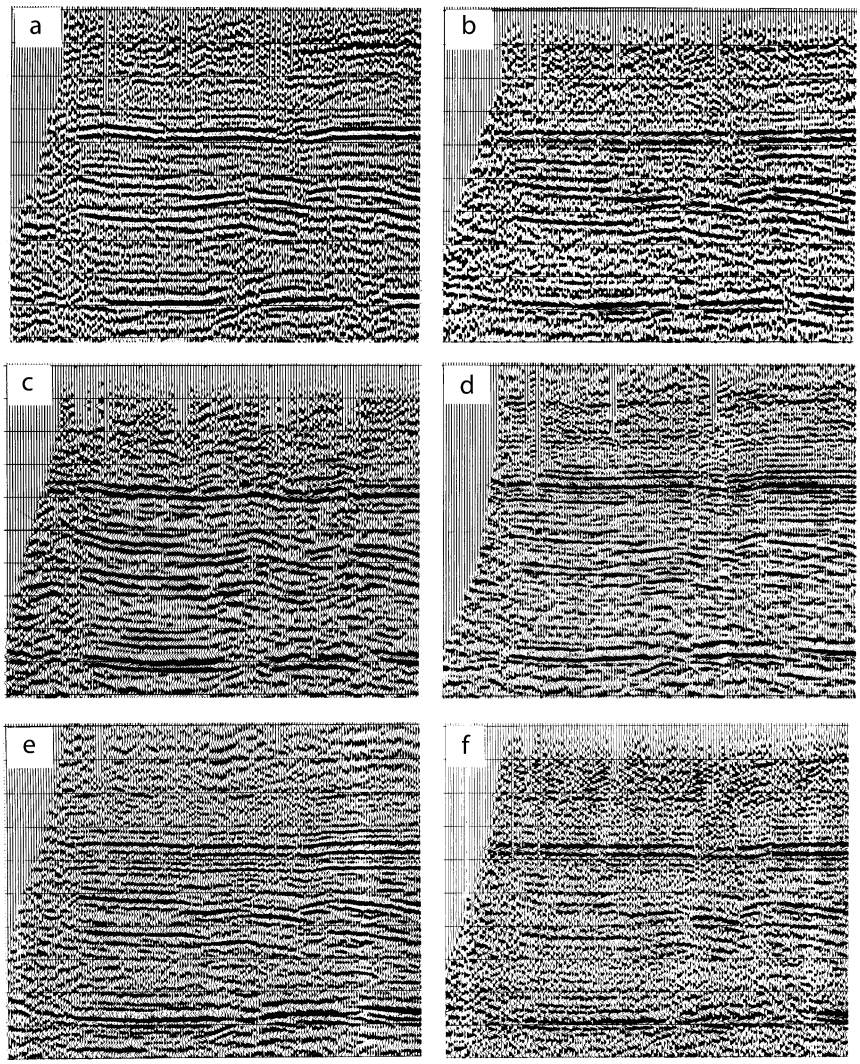
Input module → a series of processing-modules → Output module

Modules have been developed for carrying out specific tasks within the flow, e.g., static correction, band-pass filtering, stacking, migration, etc. Usually, there is a choice of modules (algorithms) available for a specific step – each with slightly different characteristics (and artifacts), and the proper selection of the modules for a flow needs both expertise and experience. This point is illustrated in the Figure 6, which shows six *different* results of processing the same data.

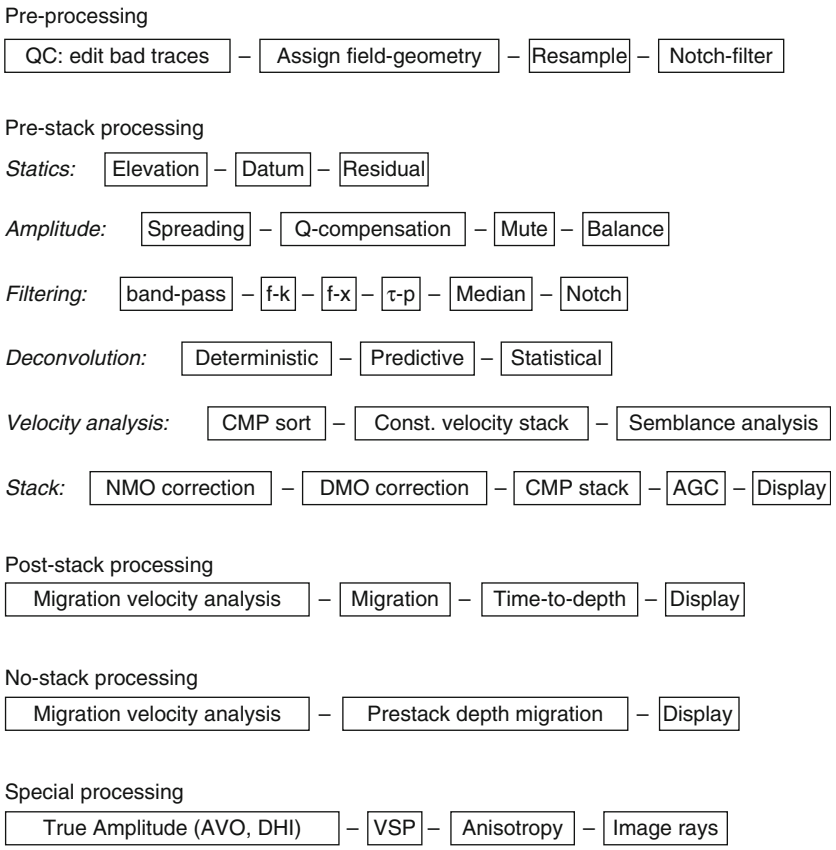
An overview of commonly applied *corrections* (processing modules) is shown in Figure 7. Space constraints will permit us to briefly describe only selected items from this list, which itself is not exhaustive; see Yilmaz (2001) for a more detailed treatment, and "*Seismic Imaging, Overview*" for additional information. Note that some modules may be applied more than once in the flow, and also, that parts of the flow may be iteratively repeated, till a reasonable result is obtained. The latter shows the importance of quality control (Q/C), by means of visual-display and other (quantitative) tools. The decision as to whether the processing of a dataset is finished depends often on the geological objectives, technical possibilities, and managerial constraints of time and money.

Pre-processing

Editing of seismic traces is an important first step, in view of the largely automated processing sequences later. *Geometry assignment* is also an essential step at this stage, and attaches acquisition information to the traces, e.g.,



Seismic Data Acquisition and Processing, Figure 6 Seismic Data Processing has no perfect answer. Seismic cross-sections produced from the same data processed by six different contractors. (Figure from Yilmaz, 2001 courtesy SEG and the author).



Seismic Data Acquisition and Processing, Figure 7 Components of seismic processing flow.

source- and receiver- coordinates. Each seismic trace is assigned a header, to store such and other information to enable efficient inter-process communication.

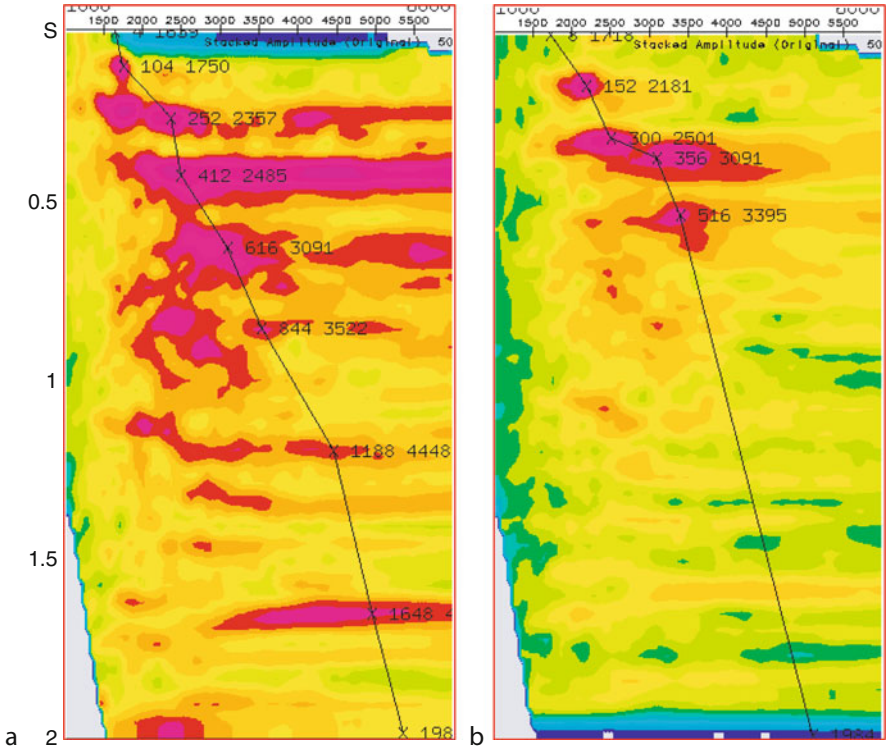
Pre-stack processing

Static corrections. These are time-invariant corrections applied to the traces, due to, e.g., elevation differences, and involve up-, or, down-shifting the entire trace in time; an example of their dramatic effect can be seen in Figure 8. The static effects due to slow lateral changes (long wavelength statics) are particularly difficult to model and can cause imaging problems. *Residual statics* involves *small* time-shifts applied late in the flow to improve the result; it uses the powerful concept of *surface-consistency* to try to correct for near-surface errors that were not modeled properly in the earlier stages. Its implementation by Rothman (1985) heralded the use of non-linear optimization (*simulated annealing, genetic algorithm*) in seismics.

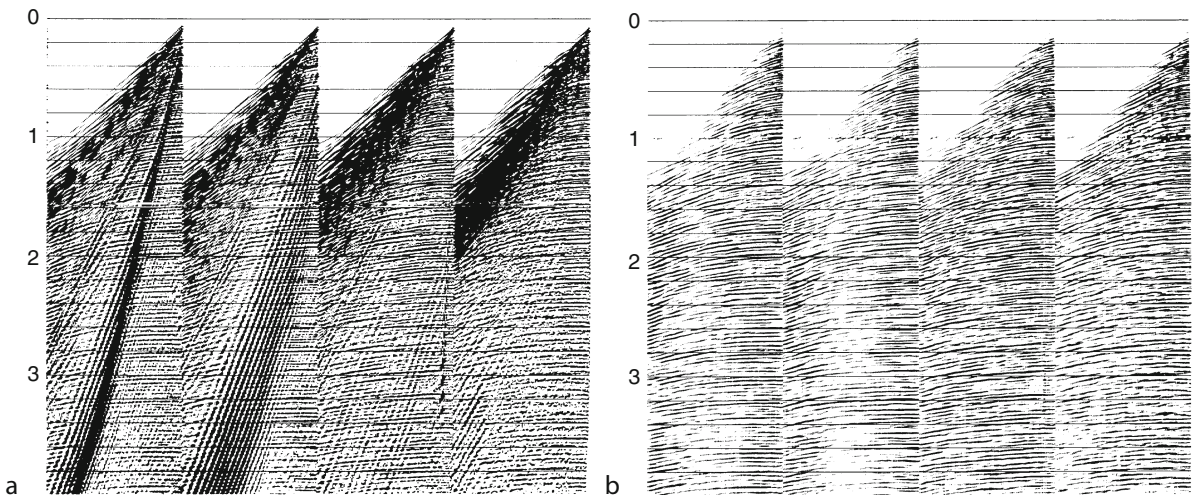
Amplitude corrections. Loss of amplitude due to *geometrical spreading* and *absorption* can be corrected for using the theory described earlier; the latter needs a Q-model, in the absence of which empirical relationships based on the total travel path/time are used. A part of the record may be removed from processing due to the presence of noise, or, suspected non-primaries; depending upon the part of the data volume removed, one then talks about *top-mute, bottom-mute, or a generalized mute*. Similarly, a *balancing* (amplitude equalization) may be applied to several adjacent traces to compensate, in an ad hoc manner, for local variations, e.g., bad receiver coupling.

Filtering, sharpness, taper. Any process that removes/reduces a part of the suspected *noise* from the data is a *filter*. Frequency-filters (high-cut, low-cut, band-pass) are the simplest examples. Data $f(t)$ is transformed using Fourier theory to its spectrum $F(\omega) = A(\omega)\ \exp^{-i\omega t}$ in the frequency domain, the amplitudes mainly corresponding to noise are zeroed-out, and the data is transformed back to the time-domain. Development of algorithms for fast and efficient Fourier transform of time-series (FFT) has caused large-scale application of digital filters.

Multi-channel data enables double-transformation of $f(x, t)$ to $F(\omega, k)$, making filtering possible based upon slopes (apparent velocities) in the $\omega - k$ plane; this is particularly effective in eliminating, e.g., slow traveling ground-roll (large amplitude surface waves), which often



Seismic Data Acquisition and Processing, Figure 8 Stacking velocity analysis using semblance (color contours). Semblance values are shown for a dataset for a range of trial velocities (*horizontal axis*), and enable interactive velocity picking as a function of TWT (*vertical axis*). The right panel shows a dramatic improvement in resolution as a result of proper static correction. (Figure from Yilmaz, 2001 courtesy SEG and the author).



Seismic Data Acquisition and Processing, Figure 9 Use of two-dimensional Fourier Transform as an apparent-velocity filter for four marine-seismic records brings out (weaker) reflections. (Figure from Yilmaz, 2001 courtesy SEG and the author).

mask the primaries. An example of such filtering is shown in Figure 9. *Notch* filters are used to remove a narrow band of frequencies, e.g., a 50-Hz noise from overhead transmission line. τ-p transforms are useful in filtering multiples, and in un-tangling far-offset data for velocity analysis, these use the Radon domain for the decomposition (Phinney et al., 1981). A few general comments apply to all filters:

– Filtering is effective only to the extent of signal-noise separation in the transformed domain.
– For any filtering, there is a trade-off between sharp cut-offs in the transform-domain and oscillatory artifacts in time-domain – and vice-versa. A compromise solution to this unavoidable problem is to apply *tapers* to smoothen the cut-off and thus minimize edge-effects.

Deconvolution. This important aspect has been dealt with in some detail in an earlier section.

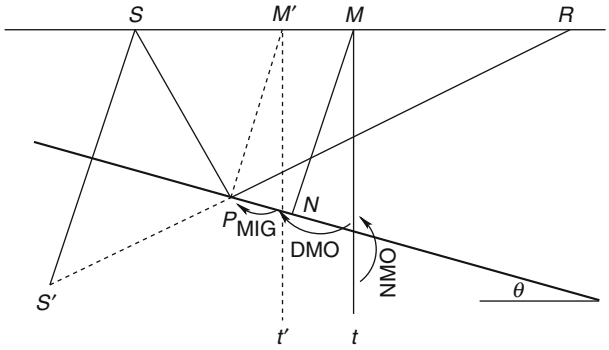
Stacking velocity analysis. This is almost always carried out in the CMP-domain, after *re-sorting* the data. The aim is to determine the velocity model, i.e., v_{rms} (TWT), to be used for computing the best move-out correction for the CMP-gathers. For a suite of velocity models, hyperbolic move-out curves are computed for the range of TWTs of interest; semblances are then computed to determine how well the arrivals in the gather line-up along these curves, and displayed in a contour plot in the v_{trial} -TWT domain, allowing interactive picking of an improved velocity model (Figure 8). The process is repeated at as many CMPs as possible – sometimes grouping neighboring CMPs together for averaging, velocity being a macroscopic property. The result is a laterally varying model of v_{rms}. Equation 6 can now be used to infer interval velocities.

CMP-stack. Once a reasonable velocity function has been determined, each trace in the CMP-gather (say, CMP_4 in Figure 1) is shifted in time by subtracting the corresponding move-out corrections. The move-out corrected traces in the CMP-gather are then added (*stacked*) together to produce one trace. This process, *CMP-stack*, reduces random noise – which does not line-up, while strengthening the reflection signal – which does, and thus improves S/N ratio of the data. Note that stacking reduces the data volume too – by a factor of *fold*! Much of the power of the seismic imaging derives from this simple step, which enhances the *primary* reflections (those only once reflected) at the expense of everything else.

Zero-offset traces/sections. The stack traces are also called *zero-offset* traces, the move-out correction having made the source and receiver coincident. A collection of stack traces is a stack- or zero-offset section, and represents the first (albeit approximate) 2-D cross-section of the subsurface. For display purposes, CMP-stack sections may be subjected to *automatic gain control* (AGC), an extremely non-linear time-variant amplitude scaling, to balance weaker/deeper signals and stronger/shallower ones.

Post-stack processing: positioning properly

The CMP-stack has one big drawback: dips were neglected throughout, which is what we are really after. This results in many artifacts in the section, e.g., crossing layering, diffraction tails, etc. Anticlinal structures are somewhat flattened, and synclinal structures could give rise to *bow-ties*.



Seismic Data Acquisition and Processing, Figure 10 Effect of dip in positioning the reflector.

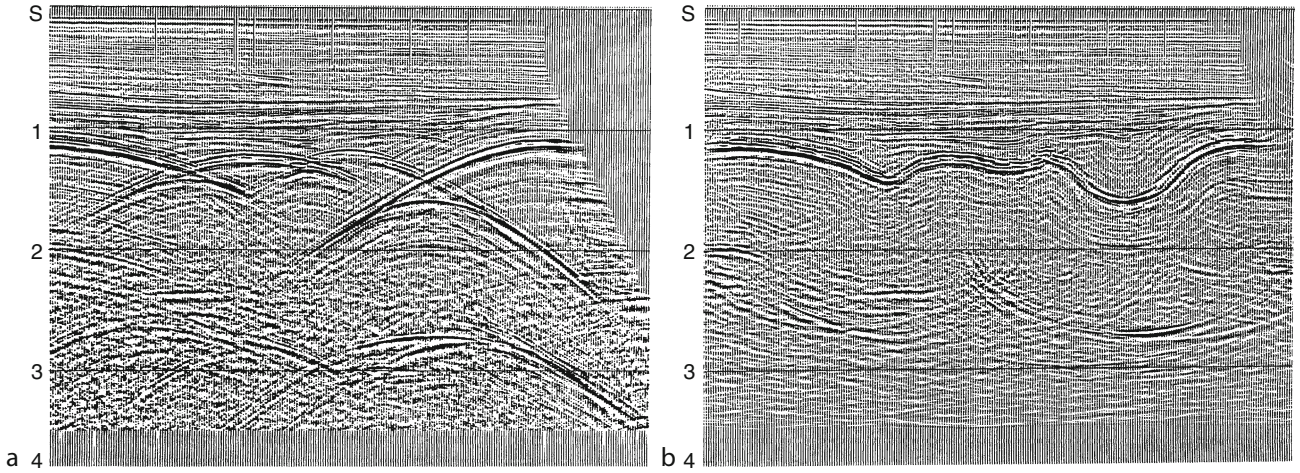
Migration. Figure 10 shows the problem schematically in the CMP-domain, for the case of a trace recorded from source S at receiver R from a reflector with a dip θ. After conventional pre-stack processing, the zero-offset trace would be plotted on the t axis below the mid-point M. This is clearly an error as the zero-offset ray for M should be incident normally on the reflector – at N; this correction is called *migration* (Yilmaz, 2001; Sheriff and Geldart, 1995).

Migration steepens and shortens energy alignments and moves these updip, clarifying the tangled image. Figure 11 shows an example of a successful migration. For further details, please see *Seismic, Migration*.

DMO. Figure 10 shows yet another error to be considered – the actual reflection point for the source-receiver combination $S-R$ is P, and not N. Even worse, for the different $S-R$ pairs making up the CMP-gather with mid-point M, the reflection points are all different, i.e., are *smeared* along the reflector, the amount of smear being dip-dependent. The process used to correct for this dip-dependent part of the move-out correction is called *DMO*. In practice, this step is applied before migration as indicated in Figure 7; Figure 10 shows the sequence:

– Reflection time is NMO corrected and plotted *below* M
– NMO corrected time is DMO corrected and plotted below M′, the *true* zero-offset point
– NMO+DMO corrected time is MIGRATED and plotted at P, the *reflection point*

Time-to-depth conversion. For final structural interpretation, the TWTs in the seismic section (stacked, or, migrated) need to be converted to depths, with velocity again playing the key role. For a homogeneous medium, this is just a rescaling of the vertical axis; with the velocity varying smoothly only in vertical direction (e.g., for flat sedimentary sequences), a nonuniform stretch of the



Seismic Data Acquisition and Processing, Figure 11 Migration positions energy from dipping structures properly. Here bow-tie like artifacts in the top part of panel (**a**) are imaged back into causative synclinal structures in panel (**b**). The artifacts persisting in the bottom of panel (**b**) probably point to lack of interest in imaging deeper structures in this case. (Figure from Yilmaz, 2001 courtesy SEG and the author).

vertical axis may suffice. Laterally varying velocities present depth-conversion problems though, increasing with the strength of the heterogeneity; ray-bending now needs to be taken into consideration.

No-stack processing: imaging complex structures
In the presence of strong lateral velocity variations (e.g., below salt structures), the conceptual model used to process CMP-gathers breaks down. Removing the simplifying assumptions makes the imaging physically more reasonable, albeit at the cost of substantially increased computational effort.

Pre-stack depth migration and migration velocity analysis. Simply put, this involves tracing the seismic energy from the source to the receiver for every recorded trace, with the philosophy that every seismic trace should be computable *if the structure and the velocity model were both known*. A detailed velocity model is essential for the success of PSDM; often a simplified model is assumed, and iteratively improved using *migration velocity analysis* (MVA). For details/issues regarding 2-D vs. 3-D, time- vs. depth- and post-stack vs. pre-stack migration, see *Seismic, Migration* and Yilmaz (2001).

Special processing
True amplitude: AVO, DHI. Observed variations of the *RC* with respect to angle of incidence may be interpreted in terms of changes in lithology across the reflecting boundary (*amplitude versus offset, or, AVO*), and may even indicate the nature of the pore-fluids. Such *direct hydrocarbon indicators* (DHI) include *bright-spots, flat-spots, polarity-reversals*, etc. (see Yilmaz, 2001; Sheriff and Geldart, 1995). A prerequisite for such analyses is *true amplitude processing*, avoiding modules that remove differential amplitude information, e.g., balancing, stacking, AGC, etc.

Converted waves. Using multi-component receivers, it is possible to identify waves that have been converted at the reflection boundary, and hence possess asymmetrical up- and down-ray-paths. Proper processing of such data, with CCP (*common conversion point*) replacing CDP, provides a better constraint for imaging.

VSP and cross-well tomography. Bore-holes can be used for placing receivers (and sources), resulting in significant noise-reduction. The first processing step now is to separate up- and down-going wave-fields, for details, see *Vertical Seismic Profiling*.

Anisotropy. Many seismic media are *anisotropic*, a common example being shales, which exhibit faster speeds parallel to the layering than across it, and require modification of procedures for proper imaging, e.g., the move-out curve would no more be hyperbolic. This field is proving important for reservoir studies too, see Helbig and Thomsen (2005) for an overview and also "Seismic Anisotropy."

Current investigations, controversies

Noise as a seismic source

This has become an exciting and active field of research (*seismic interferometry and daylight imaging*), connecting early conjectures (Claerbout, 1968) and more recent breakthroughs (Fink, 1993, 1997). See *Seismic Noise*; *Seismic, Ambient Noise Correlation* and Schuster et al. (2004) for an introduction to this evolving area.

Imaging versus inversion

Imaging tries to obtain useful (drillable) structural information using large data-redundancy and simple conceptual models, whereas inversion aims at getting values for the physical parameters of the medium, using more involved theory. *Full wave-form inversion* (see *Seismic, Waveform Modeling and Tomography*) i.e., predicting each seismic/seismological trace completely in terms of the viscoelastic properties of the medium is the ultimate goal. Note, that *perfect inversion implies perfect imaging – and vice-versa!*

Summary

The simple *echo-in-the-well* experiment mentioned at the start needs many physico-mathematical supports when applied to the earth's subsurface. Starting at data acquisition, the modules yielding the final image resemble a pipeline (flow). Several of these have been explained briefly; for others, cross-references elsewhere in this volume have been provided.

Bibliography

Aki, K., and Richards, P. G., 2002. *Quantitative Seismology*, 2nd edn. Sausalito: University Science Books.

Backus, M., 1959. Water reverberations – their nature and elimination. *Geophysics*, **24**(2), 233–261.

Claerbout, J. F., 1968. Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, **33**, 264.

Claerbout, J. F., 1985a. *Fundamentals of Geophysical Data Processing*. Oxford: Blackwell. URL: http://sepwww.standford.edu/sep/prof/fgdp5.pdf.

Claerbout, J. F., 1985b. *Imaging the Earth's Interior*. Oxford: Blackwell. URL: http://sepwww.stanford.edu/sep/prof/iei2/.

Dennison, A. T., 1953. The design of electromagnetic geophones. *Geophysical Prospecting*, **1**, 3–28.

Dix, C., 1955. Seismic velocities from surface measurements. *Geophysics*, **20**, 68–86.

Fink, M., 1993. Time reversal mirrors. *Journal of Physics. D. Applied Physics*, **26**, 1333–1350. URL: http://iopscience.iop.org/0022-3727/26/9/001.

Fink, M., 1997. Time reversed acoustics. *Physics Today*, **50**, 34–40. URL: http://dx.doi.org/10.1063/1.881692.

Helbig, K., and Thomsen, L., 2005. 75-plus years of anisotropy in exploration and reservoir seismics: A historical review of concepts and methods. *Geophysics*, **70**, 9ND–23ND.

Lee, M., 2008. *Comparison of the modified Biot-Gassmann theory and the Kuster-Toksöz theory in predicting elastic velocities of sediments*. U.S. Geological Survey Scientific Investigations Report.

Liner, C., 2004. *Elements of 3D Seismology*. Tulsa: PennWell.

Menke, W., 1989. *Geophysical Data Analysis*, rev. edn. Academic Press, New York, USA.

Phinney, R., Roy Chowdhury, K., and Frazer, L. N., 1981. Transformation and analysis of record sections. *Journal of Geophysical Research*, **86**(B1), 359–377.

Robinson, E., 1967. Predictive decomposition of time series with application to seismic exploration. *Geophysics*, **32**, 418–484.

Robinson, E., 2005. The MIT Geophysical Analysis Group (GAG): 1954 and beyond. *Geophysics*, **70**, 7JA. URL: http://dx.doi.org/10.1190/1.2000287.

Robinson, E., and Treitel, S., 1964. Principles of digital filtering. *Geophysics*, **29**, 395–404.

Rothman, D., 1985. Nonlinear inversion, statistical mechanics, and residual statics estimation. *Geophysics*, **50**(12), 2784–2796.

Schuster, G. T., Yu, J., Sheng, J., and Rickett, J., 2004. Interferometric/daylight seismic imaging. *Geophysical Journal International*, **157**, 838–852.

Sheriff, R. E., and Geldart, L. P., 1995. *Exploration Seismology*, 2nd edn. Cambridge, UK: Cambridge University Press.

Treitel, S., 2005. The MIT Geophysical Analysis Group (GAG): 1954 and beyond. *Geophysics*, **70**, 31JA. URL: http://dx.doi.org/10.1190/1.1993707.

Vermeer, G. J., 2002. *3-D Seismic Survey Design*. Tulsa: Society of Exploration Geophysicists.

Yilmaz, O., 2001. *Seismic Data Analysis, Processing, Inversion and Interpretation of Seismic Data*, 2nd edn. Tulsa: Society of Exploration Geophysicists. Investigations in Geophysics, Vol. 1.

Cross-references

Energy Partitioning of Seismic Waves
Inverse Theory, Global Optimization
Ocean Bottom Seismics
Propagation of Elastic Waves: Fundamentals
Seismic Anisotropy
Seismic Diffraction
Seismic Imaging, Overview
Seismic Instrumentation
Seismic Noise
Seismic Properties of Rocks
Seismic Waves, Scattering
Seismic, Ambient Noise Correlation
Seismic, Migration
Seismic, Ray Theory
Seismic, Reflectivity Method
Seismic, Viscoelastic Attenuation
Seismic, Waveform Modeling and Tomography
Single and Multichannel Seismics
Vertical Seismic Profiling

SEISMIC DIFFRACTION

Enru Liu
China University of Mining and Technology, Xuzhou, China

Definition

Diffraction – Redistribution in space of the intensity of waves resulting from the presence of an object. It is also referred as the penetration of wave energy into areas forbidden by geometrical optics, e.g., the bending of wave energy around obstacles without obeying Snell's law as explained in Huygens' principle (generation of secondary sources).

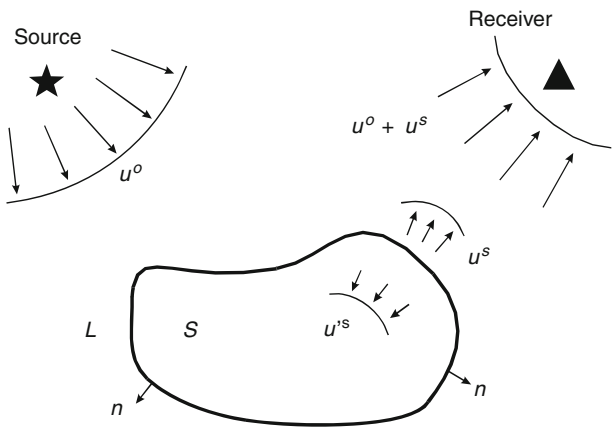
Diffraction wave-field – An event observed on seismic data produced by diffracted energy, resulting at the termination of reflectors (as at faults and other abrupt changes in seismic impedance), and it is characterized on seismic records and sections by a distinctive alignment.

Diffraction tomography – An inverse technique that is used in seismic exploration to reconstruct the physical properties under investigation using wave-equation propagation.

Introduction

When a wave (elastic wave, electromagnetic wave, or sound wave) meets an object, particle, or obstacle, it is diffracted due to scattering of energy of the propagating wave (Figure 1). However, in the literature, the terms of diffraction and scattering are often used interchangeably, and it can be sometimes confusing. Scattering and diffraction are two physical phenomena that any kind of waves can experience, but they are not the same thing. Scattering is understood in terms of particles, and behaves similarly for waves. Scattering is effectively bouncing off something. For waves, it is being absorbed and then almost immediately released in another direction. Scattering occurs because an object gets in the way of the wave. The part of the wave that strikes the object must either pass through (e.g., light through glass), be absorbed (sunburn), or be scattered (light bouncing off the wall, so we can see the wall). Diffraction is due to part of a wave being removed. It is an action taken by the part of the wave that does not strike an object. Here is an example of diffraction: Imagine a straight wave traveling forward along the surface of water. If you block the left half of the wave, the right half will not just keep moving forward, and it will expand toward the left, toward where the blocked wave would have been. A wave seems to continuously regenerate itself, always pushing forward on itself. When a section is removed, parts of the wave get pushed into the empty spot. This, in some ways, correlates to your pushing a wide row of blocks. If many people push on a few blocks each, a straight line can be maintained. If one person tries to do so, the blocks in front will tend to spread out.

In a sense, diffraction and scattering refer to a wave being redirected as a result of interacting with objects. However, a more precise definition used in optics distinguishes these two. In optics, scattering generally implies interaction of waves (or photons) with spatially uncoordinated (unordered) atoms, i.e., wave being scattered off the particles and surfaces. This means that if one looks at a picture of the scattered radiation, one would see a spatially continuous footprint. Diffraction, on the other hand, occurs when the object or part of the object is made up of ordered atoms. These atoms, being neatly arranged, "scatter" the waves or photons in a coordinated way, i.e., in specific directions, giving rise to what we can see on a film as bright spots rather. In other words, diffraction is a special type of scattering that leads to large-scale interference effects. Usually this is because the surface causing the scattering has some overall organization, such as a ruled diffraction grating or the knife-edge of a slit.



Seismic Diffraction, Figure 1 Problem configuration: a scattering object S bounded by the curve L with outward normal n. Upon an incidence of u_0 located at source, the total wave-field received at receiver is the superposition of the incident wave-field u_0 and the scattered wave-field u_s.

Although scattering and diffraction are not logically separate, the treatments tend to be separated, with diffraction being associated with departures from geometrical optics caused by the finite wavelength of the waves. Thus, diffraction traditionally involves apertures or obstacles whose dimensions are large compared to a wavelength. To lowest approximation the interaction of waves is described by ray tracing (geometrical optics). The next approximation involves the diffraction of the waves around the obstacles or through the apertures with a consequent spreading of the waves. Simple arguments show that the angle of diffraction of the waves are confined to the region $\theta \leq \lambda/a$ where λ is the wavelength and a is linear dimension of the aperture or obstacle (approximations considered work well if $\lambda/a \leq 1$). Note that diffraction may cause the localization of seismic waves – a phenomenon that is similar to the localization of lights in crystal (Larose et al., 2004). This phenomenon is caused by the focusing and defocusing of energy when seismic wave propagates through media with distributions of periodical or random distribution of scattering bodies.

In order to gain some perspective on these two extremely complex phenomena (diffraction and scattering), various theories and models have been developed in physics, and the limitations and validity of these theories are controlled by two ratios: object dimension α to wavelength (α/λ) and path-length L to wavelength λ (L/λ). In contrast to other branches of physics, in geophysics path-length is also important as we are interested both in near field as well as far field (and often near and far fields are treated differently). We also often use the dimensionless parameters $k\alpha$ and kL (where $k = \omega/\lambda$ is the wave number, ω is frequency).

Diffraction theories

Application of any diffraction model can be divided into two separate tasks. First, one must obtain the fields exiting a diffracting object (i.e., the near fields, or the boundary field values), or a reasonable approximation thereof. The second step involves propagating those fields to the

desired observation point. These are distinct and separate parts of the diffraction models. Most texts do not make this separation clear. Instead, the boundary value assumptions and subsequent propagation into the far field are lumped together into one theoretical treatment. If the resulting diffraction pattern is at all inaccurate, it is difficult to determine how much of that error is due to incorrect boundary fields and how much is the result of the propagation calculation. Because of this, it is often difficult to know which model is appropriate for a particular problem.

There are a number of different models to compute diffraction wave-field due to wave scattering, including the Huygens' Principle, the Rayleigh-Sommerfeld theory, the Kirchhoff's diffraction theory, Taylor series perturbation theory or the high-order Born approximation, Rytov phase approximation, and a model referred to as angular spectrum of plane waves. The well-known Fraunhofer and Fresnel approximations, as they appear in most introductory texts, are derived from the Kirchhoff model. Several other methods are also available for treating diffraction problems: discrete wave-number techniques; generalized ray techniques; and various numerical methods (finite difference methods; finite element method; and boundary integral or element methods). Each of these theoretical methods and models is based on some assumptions and has its strengths and weaknesses, and each can be satisfactorily employed for some ranges of problems. The choice of an appropriate model is based on what is known about a specific problem.

If there are several objects or particles, the scattering from one object will induce further scattered fields from all the other objects, which will induce further scattered fields, from all the other objects, and so on. This process is called multiple scattering. Not all theories are applicable to multiple scattering problems (in physics, this is called many-body problem). The Kirchhoff approximation ignores multiple scatterings between any two surface points. In general, it has been considered valid for the large-scale objects. Perturbation theory based on the Taylor series expansion, also sometimes called the high-order Born approximation, is valid for the small-scale objects whose dimensions are less than a wavelength (or objects whose physical properties are not too different from background solids). The Rytov phase approximation to large-scale object is not subject to the stringent restrictions that apply to the Kirchhoff approximation. Studies have shown that the Rytov approximation improves the Kirchhoff approximation in both amplitude and phase. To some degrees, the high-order Born series approximation can account for multiple scattering between closely-spaced objects. For instance, the second-order Born approximation might be sufficient to guarantee the accuracy for general rough surfaces without infinite gradients and extremely large surface heights. In contrast to other branch of field, e.g., optics, in seismology two kinds of waves exist, compressional and shear waves. These two waves can convert to each other when one meets an object. When multiple objects exist, the conversion and interaction between different wave types due to multiple scattering can be very complex. Therefore, care must be taken when one uses any diffraction theory to solve specific geophysical problems.

Geometrical theory of diffraction

The geometrical theory of diffraction (GTD) is an extension of geometrical optics that accounts for wave diffraction by edges. It was introduced in 1953 by Keller (the most commonly cited reference was published by Keller in 1962). The geometrical theory of diffraction was devised to eliminate many of the problems associated with geometrical optics. The strongest diffracted fields arise from edges, but ones of lesser strength originate from point discontinuities (tips and corners). The total field $u = (u_1, u_2, u_3)$ at an observation point $\vec{x}$ is decomposed into geometrical optic rays (the incident or reference field) u_i^0 and diffracted components u_i^d

$$u_i(\vec{x}) = u_i^0(\vec{x}) + u_i^d(\vec{x}). \tag{1}$$

The behavior of the diffracted field is based on the following postulates of GTD:

1. Wavefronts are locally plane waves.
2. Diffracted rays emerge radially from an edge.
3. Rays travel in straight lines in a homogeneous medium.
4. Polarization is constant along a ray in an isotropic medium.
5. The diffracted field strength is inversely proportional to the cross sectional area of the flux tube.
6. The diffracted field is linearly related to the incident field at the diffraction point by a diffraction coefficient (see Achenbach et al., 1982, for various analytic solutions).

GTD is a high frequency method for solving wave scattering problems from large-scale discontinuities or discontinuities in more than one dimension at the same point, and it uses ray diffraction to determine diffraction coefficients for each diffracting object-source combination. These coefficients are then used to calculate the field strength and phase for each direction away from the diffracting point. These fields are then added to the incident fields and reflected fields to obtain a total solution. Multiple scattering wave-fields cannot be easily computed using GTD.

Kirchhoff approximation

In the Kirchhoff representation of diffracted wave-fields, the ith component of diffracted wave-field u_i is computed using

$$u_i^d(\vec{x}) = -\int_{\sum} [u_k(\vec{x})] c_{kjpq}^0 \left[\frac{\partial G_i^p(\vec{x},\vec{X})}{\partial x_q}\right] n_j dS_{\vec{x}}, \tag{2}$$

where n_j is the jth component of the normal n to the surface of scattering object, $\mathbf{X}$ is a point on the face of the scattering object, $[u_k]$ is the kth displacement discontinuity across the object in the direction of n (object normal), and c_{kjpq}^0 is the elastic tensor of the background, which are often assumed to be isotropic. Equation 2 provides

a means of evaluating the diffracted field so long as the displacement discontinuity $[u]$ on the object can be estimated accurately. Since the amplitudes and phases of $[u]$ are unknown, in the Kirchhoff approximation, these are taken to be the same as if the object were infinitely long, that is, the effect of the boundary is ignored (see Douglas and Hudson, 1990; Liu et al., 1997). Therefore, the Kirchhoff approximation is a high frequency approximation, which is only applicable to diffraction from objects whose dimension is larger than the wavelength.

Perturbation theory: Born and Rytov approximation

The diffracted wave-field in Equation 1 can be written as an infinite series of the Taylor series expansion and is derived by assuming that the physical property of scattering objects is written as a perturbation of background media. The Born approximation consists of taking the incident field in place of the total field as the driving field at each point in the scatterer. It is the perturbation method applied to scattering by an extended body. It is accurate if the scattered field is small, compared to the incident field, in the scatterer. It is only valid for weak scattering or when the obstacles are small compared to both the wavelength and the spacing between the objects. Clearly, it has serious limitations when dealing with large-scale objects. The simplest approximation, called single scattering or the first-order Born approximation, is to ignore the multiply scattered field between objects. This approximation has been widely used in geophysics (Hudson and Heritage, 1981; Wu, 1982, 1989).

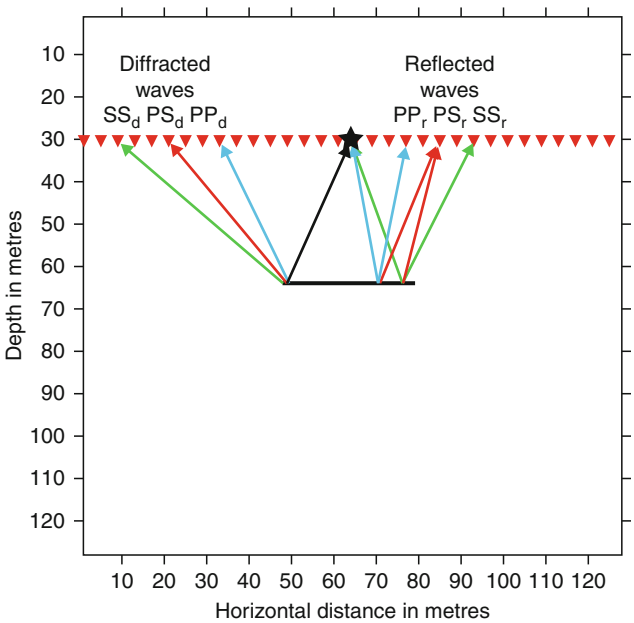
The Rytov approximation linearly perturbs the phase of a wave-field with respect to model parameters such as velocity whereas the Born approximation perturbs the amplitude. When the Green's functions for point sources are replaced by Eikonal approximations, the Rytov perturbed wave-field becomes a scaled, differentiated, time-delayed version of the reference or incident wave-field.

Numerical methods used to compute diffraction wavefield

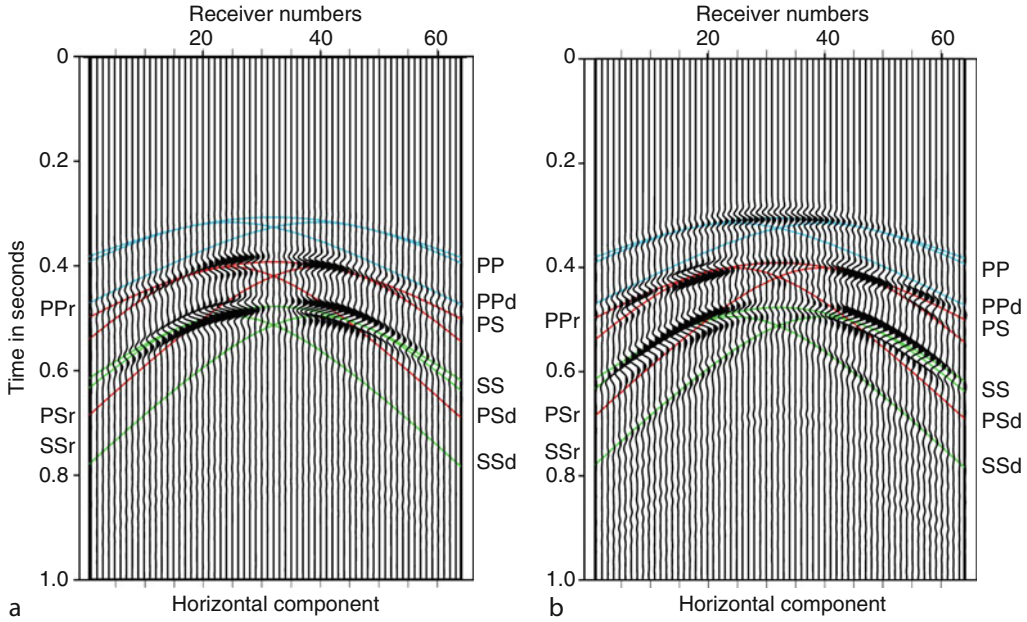
Various numerical methods can also be used to compute diffracted wave-fields, particularly now, when high-performance computers are widely available. In geophysics, finite difference methods have been used widely in the study of scattering of elastic waves by crustal heterogeneities with continuous variation of physical properties and they have also been used to model scattering by thin cracks and large fractures (see Coutant, 1989; Fehler and Aki, 1978; Coates and Schoenberg, 1995; Vlastos et al., 2003, 2007). Elastodynamic boundary integral equation or boundary element method has also been widely used to compute wave-fields from discrete inclusions with various spatial distributions as well as rough-surface topographic variations (see Bouchon, 1987; Pointer et al., 1998; Liu and Zhang, 2001; Sanchez-Sesma and Campillo, 1991, 1993).

An example of seismic diffraction from a single fracture

Some practical applications of diffraction theories in seismology include scattering from cavities; topographic variation, fractures, cracks (Figure 2). Here we give three examples to demonstrate the application of various theories in tacking diffraction problems. An example is given here for diffraction from a single fracture as computed using the finite difference method (Vlastos et al., 2003, 2007). The model geometry is shown in Figure 2. The source, receivers, and fracture are situated in an ideal elastic full space (Vp = 3,300 m/s, Vs = 1,800 m/s, density $\rho = 2.2$ g/m^3). The receiver array at which vertical and horizontal particle displacements are recorded is horizontal and 340 m above the fracture. The fracture is 300 m long. The source is located at the center of the receiver array. The source type is a vertical force. The source signal is a Ricker wavelet with a peak frequency of 25 Hz and a pulse initial time of 0.1 s. Figure 3 also shows the different kinds of waves generated by the interaction of the waves generated by the source and the fracture. The source generates both P and S waves. When they reach the fracture boundary those waves are reflected and we have PPr, PSr, SPr, and SSr waves. We calculate the theoretical ray travel-times and overlap them on the synthetic seismograms. Figures 3a and 3b show the horizontal (x) and the vertical (z) components, respectively, of the synthetic seismograms together with the theoretical ray travel-times. As we can see from both figures, we have very good agreement between the theoretical ray



Seismic Diffraction, Figure 2 Schematic representation showing diffraction from a fracture, and representation of the ray paths of the different kind of waves generated by the source that interact with the fracture.



Seismic Diffraction, Figure 3 Comparison between the theoretical ray travel-times and the synthetic seismograms generated by the modeling method: (**a**) the horizontal (*x*) and (**b**) the vertical (*z*) components.

travel-times and the synthetic seismograms. All types of waves are accurately represented in the synthetic seismograms. Owing to the type of source that we implement, we have strong arrivals at short offsets on the horizontal component and strong arrivals at long offsets on the vertical component. In addition to that, the diffracted waves from the tips of the fracture and the PPr and PPd waves are not visible in the horizontal component, but they are very clearly demonstrated in the vertical component and follow the theoretical travel-times. This is expected because the source causes vertical displacements on the medium, so very close to the source and very far away from it, the horizontal displacement is negligible. Another aspect of the comparison between the theoretical and the modeled data is that they give us further insight into the waveform patterns. For instance, we can see in both Figures 3a and 3b that in the areas of superposition between the reflected waves from the fractures and the diffracted waves from the tips we have maximum amplitude in the wave-field, as a result of constructive interference. This gives us valuable information concerning the medium we are examining.

Diffraction tomography

Seismic tomography is emerging as an imaging method for determining subsurface structure. When the view-angle coverage is limited and the scale of the medium inhomogeneities is comparable with the wavelength, as is often true in geophysical applications, the performance of ordinary ray tomography becomes poor. Other tomographic methods are needed to improve the imaging process, e.g., diffraction tomography. It has been widely used in surface reflection profiling (SRP), vertical seismic profiling (VSP), and cross-hole measurements. Theoretical formulations are derived by Wu and Toksoz (1987) for two-dimensional geometry in terms of line sources along a source line and line receivers along a receiver line. The theory for diffraction tomography is based on the Born or Rytov approximation. Multisource holography, which is similar to Kirchhoff-type migration, often gives distorted images of the object. This distortion causes long tails of the image in the case of SRP and a strong noise belt in the case of VSP and is due to incomplete and nonuniform coverage of the object spectrum. The filtering operation of diffraction tomography helps in correcting the nonuniform coverage (including duplication) of the object spectrum in the reconstruction process and therefore reduces the distortions. On the other hand, multisource holography is better suited for imaging sharp boundaries with large acoustic impedance contrasts since diffraction tomography is restricted, to weak inhomogeneities. In addition, multisource holography has the flexibility to be used with an arbitrary number of sources (including a single source). Its sampling interval is not restricted by the Nyquist frequency. Numerical examples show that combined data sets (such as surface reflection data combined with VSP data or cross-hole data combined with surface data) improve the image quality.

Summary

Diffraction refers to the spatial distribution of the intensity of seismic waves resulting from the presence of an object

(e.g., a hill- or valley-like topographic feature on the surface, or a fracture, cavity, or cave in the subsurface). It is also referred as the penetration of wave energy into areas forbidden by geometrical optics, e.g., the bending of wave energy around obstacles without obeying Snell's law as explained in Huygens' principle (as secondary sources). In the geophysical literature, the words of diffraction and scattering are often used interchangeably and it can be confusing. Diffraction and scattering are two different physical phenomena, but they are related to each other. Several analytic diffraction theories have been developed, e.g., geometrical theory of diffraction and Kirchhoff diffraction theory. More recently, numerical methods, such as finite difference and boundary element or boundary integral methods, are becoming increasingly used by geophysicists to simulate wave diffractions by complex variation of Earth's topography or subsurface cavies (cavities), fractures, irregular layers, etc. Geophysicists now often use the diffracted wave-field to reconstruct the subsurface physical properties (diffraction tomography) to solve the so-called inverse diffraction problem.

Bibliography

Achenbach, J. D., Gautesen, A. K., and McMaken, H., 1982. *Ray Methods for Waves in Elastic Solids, with Application to Scattering by Cracks*. London: Pitman Learning.

Bouchon, M., 1987. Diffraction of elastic waves by cracks or cavities using the discrete wave-number method. *The Journal of the Acoustical Society of America*, **81**, 1671.

Coates, R. T., and Schoenberg, M., 1995. Finite-difference modelling of faults and fractures. *Geophysics*, **60**, 1514.

Coutant, O., 1989. Numerical study of the diffraction of elastic waves by fluid-filled cracks. *Journal of Geophysical Research*, **94**, 17805.

Douglas, A., and Hudson, J. A., 1990. The effect on teleseismic P of the zone of damage created by an explosion. *Geophysical Journal International*, **103**, 111.

Fehler, M., and Aki, K., 1978. Numerical study of diffraction of plane elastic waves by a finite crack with application to location of a magma lens. *Bulletin of the Seismological Society of America*, **68**, 573.

Hudson, J. A., and Heritage, J. R., 1981. The use of the Born approximation in seismic scattering problems. *Geophysical Journal International*, **66**, 221.

Keller, J. B., 1962. Geometrical theory of diffraction. *Journal of the Optical Society of America*, **52**, 116.

Larose, E., Margerin, L., van Tiggelen, B. A., and Campillo, M., 2004. Weak localization of seismic waves. *Physical Review Letters*, **93**, 048501-1–048501-4.

Liu, E., and Zhang, Z. J., 2001. Numerical study of elastic wave scattering by distributed cracks or cavities using the boundary integral method. *Journal of Computational Acoustics*, **9**, 1039.

Liu, E., Crampin, S., and Hudson, J. A., 1997. Diffraction of seismic waves by cracks with application to hydraulic fracturing. *Geophysics*, **62**, 253.

Pointer, T., Liu, E., and Hudson, J. A., 1998. Numerical modelling of seismic waves scattered by hydrofractures: application of the indirect boundary element method. *Geophysical Journal International*, **135**, 289.

Sanchez-Seama, F. J., and Campillo, M., 1993. Topographic effects for incident P, SV, and Rayleigh waves. *Tectonophysics*, **218**, 113.

Sanchez-Sesma, F. J., and Campillo, M., 1991. Diffraction of P, SV, and Rayleigh waves by topographic features: a boundary integral formulation. *Bulletin of the Seismological Society of America*, **81**, 2234.

Vlastos, S., Liu, E., Main, I. G., and Li, X. Y., 2003. Numerical simulation of wave propagation in media with discrete distributions of fractures: effects of fracture sizes and spatial distributions. *Geophysical Journal International*, **152**, 649.

Vlastos, S., Liu, E., Main, I. G., and Narteau, C., 2007. Numerical simulation of wave propagation in fractured media: scattering attenuation at different stages of the growth of a fracture population. *Geophysical Journal International*, **171**, 865.

Wu, R. S., 1982. Attenuation of short period seismic waves due to scattering. *Geophysical Research Letters*, **9**, 9.

Wu, R. S., 1989. The perturbation method in elastic wave scattering. *Pure and Applied Geophysics*, **131**, 605.

Wu, R. S., and Toksoz, M. N., 1987. Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, **52**, 11.

Cross-references

Deep Seismic Reflection and Refraction Profiling
Earthquake Rupture: Inverse Problem
Propagation of Elastic Waves: Fundamentals
Seismic, Migration
Seismic, Ray Theory
Seismic Waves, Scattering

SEISMIC DISCONTINUITIES IN THE TRANSITION ZONE

Lev P. Vinnik
Institute of Physics of the Earth, Moscow, Russia

Definition

The transition zone (TZ) is the mantle layer bounded by the 410– and 660–km seismic boundaries. The high P-wave and S-wave velocity gradients within the TZ are caused by a series of polymorphic phase transitions, the depths (pressures) of which are controlled by temperature and composition. Structure of the TZ plays an important role in the heat/mass transfer between the upper and the lower mantle.

Mineral physics data on the phase transitions in the TZ

The most frequently used model of mantle composition is pyrolite which contains ~60% of olivine $(Mg,Fe)_2SiO_4$. At a depth of ~410 km olivine (α) transforms to wadsleyite (β, modified spinel). The Clapeyron slope of this transition is positive (4.0 MPa/K, Katsura et al., 2004); the increase of the S-wave velocity is ~12%. At a depth of ~520 km wadsleyite transforms to ringwoodite (γ, silicate spinel). The velocity increase at this transition is by an order of magnitude less than at the α/β transition (e.g., Ita and Stixrude, 1992; Cammarano et al., 2005). At a depth of ~660 km ringwoodite transforms to a mixture of perovskite $(Mg,Fe)SiO_3$ and

magnesiowüstite (Mg,Fe)O. The S velocity contrast of this transition is comparable to that of the α/β transition. The post-spinel transition is sharp and has a negative Clapeyron slope of −2.8 MPa/K (Hirose, 2002). This value is disputed and some recent estimates are in a range from −0.4 to −1 MPa/K (Ohtani and Sakai, 2008).

The other components of pyrolite are orthopyroxene, clinopyroxene, and garnet. These components experience in the TZ more gradual transformations: the pyroxenes dissolve into garnet (majorite) and majorite transforms to perovskite near the bottom of the TZ. The Clapeyron slope of the post-majorite transition is 1.3 MPa/K (Hirose, 2002). The post-spinel and post-majorite phase boundaries cross each other at 1,700–1,800°C, and the corresponding seismic discontinuity is a combined effect of both transitions. At the temperatures less than 1,700°C the discontinuity is formed mainly by the post-spinel transition, whilst at the higher temperatures the post-majorite transition becomes dominant (Hirose, 2002).

Seismic methods

The actual composition and temperature of the TZ can be constrained by seismic data. An increase of the seismic velocity gradient with depth is mirrored by the increase of the apparent velocity of the P and S seismic arrivals with distance. Details can be obtained from the analysis of the related loop (triplication) in the travel times and modeling the related waveforms with synthetic seismograms (e.g., Grand and Helmberger, 1984). First indications of an increase of the velocity gradient near the top of the TZ were obtained from the analysis of incidence angles of the P waves (Galitzin, 1917). The 660-km discontinuity at the bottom of the TZ with a comparable velocity contrast was discovered in the mid-1960s (Anderson, 1965). Prior to the discovery of the 660-km boundary, the TZ was defined as the region between the 400- and 1,000-km depths (Bullen's region C).

The triplications corresponding to discontinuities in the TZ are observed at epicentral distances between ∼1,700 and ∼3,200 km, whereas properties of the mantle may change significantly on a scale of a few hundred kilometers. Observations of reflected and mode-converted phases can provide better resolution. In the neighboring field of exploration seismology, accurate mapping of discontinuities is based on observations of reflected phases. These signals are small relative to noise, and the detection is performed by using receiver arrays and stacking the recordings of different sensors with move-out time corrections. These methods are very efficient, but in seismology adequate receiver arrays are few. The idea of receiver function approach in seismology is to replace receiver arrays by seismic-source arrays that can be very large and dense. The differences in the waveforms of individual earthquakes that are recorded at the same station are eliminated by appropriate frequency filtering.

P-wave receiver functions (PRFs) present the oldest and most usable variety of receiver functions (Vinnik, 1977). This technique is based mainly on observations of Ps phases, converted from P to S. The delay of the Ps phase relative to P depends on the depth of the discontinuity and velocities along the wave-propagation paths. The amplitude of the Ps phase is proportional to the S velocity contrast at the discontinuity. A gradual transition is transparent for short-period P waves, and a shift of the spectrum of the Ps phase to lower frequencies indicates the width of the discontinuity. The detection of small seismic phases in PRFs can be accompanied by measurements of their slowness. In a laterally homogeneous Earth, Ps phases from the TZ discontinuities differ by slowness from lithospheric reverberations that arrive in the same time window, and this difference may help to separate the signals from noise. The best results can be obtained by combining source and receiver arrays. Unfortunately, practical detection of the TZ converted phases in noise is often based on relaxed criteria. Some of the widely cited results are obtained from noisy receiver functions which show a continuous train of positive "swells" and negative "troughs." The "swells" are brightly painted and the resulting optical illusion may convince readers that they see separate arrivals rather than an interference pattern.

S-wave receiver functions (SRFs) (Farra and Vinnik, 2000) are complementary to PRFs and deal with the Sp phases converted from S to P. Multiple scattering at shallow discontinuities, which presents the major source of noise in PRFs, is practically absent in SRFs, because the Sp phases from large depths arrive much earlier than the scattered phases from shallow discontinuities. Another useful method is based on observations of ScS reverberations (Revenaugh and Jordan, 1991). Detection of the phases reflected from TZ discontinuities in this method is in principle similar to that employed in receiver functions and this technique can be viewed as a variety of receiver function techniques. A shortcoming of this method is its low (in comparison to PRFs and SRFs) lateral resolution. Precursors to the seismic phases SS (SH component), PP and P′P′ (SH, P and P′ waves reflected from the Earth's surface between the source and the receiver) include phases reflected from TZ discontinuities. These phases can be detected by using the receiver function approach (Shearer, 1991). Lateral resolution of the SS-precursor technique (∼1,000 km) is by an order of magnitude lower than of PRFs, but the SS precursors are useful in the studies of the TZ of remote regions, where seismograph stations are too few for high-resolution studies.

Topography and sharpness of TZ discontinuities

The amplitudes of seismic phases converted or reflected from the major TZ discontinuities at 410- and 660-km depths vary laterally but on the average are approximately two times lower than predicted for olivine. This relationship implies that the actual composition of the TZ is broadly similar to pyrolite which contains ∼60% of olivine. A comparison of synthetic and actual wave fields for the TZ indicates that the potential temperature in the mantle differs for a homogeneous pyrolite mantle and

a mechanical mixture of basaltic and olivine components: 1,720 and 1,625 K, respectively (Ritsema et al., 2009).

The amplitudes of reflected and converted phases from the 520-km discontinuity in pyrolite should be about an order of magnitude lower than from the 660-km discontinuity and, for this reason, practically unobservable. However, there are reports on observations of a discontinuity near this depth, especially in the data on SS precursors (e.g., Flanagan and Shearer, 1998; Gu et al., 1998; Deuss and Woodhouse, 2001) and P wave recordings (e.g., Ryberg et al., 1997). Revenaugh and Jordan (1991) argued that the large 520-km discontinuity in seismic data is related to garnet/post-garnet transformation. Gu et al. (1998) discussed a possibility of very deep continental roots extending into the TZ. Deuss and Woodhouse reported splitting of the 520-km discontinuity into two discontinuities. Bock (1994) denied evidence for the 520-km discontinuity in precursors to SS.

The spectra of P410s in high-quality PRFs are usually close to the spectra of P, which indicates that the 410-km discontinuity is sharp, with a width of several kilometers (Petersen et al., 1993). Benz and Vidale (1993) observed high-frequency precursors to P′P′ with implication that the widths of both 410- and 660-km transitions are ~4 km or less. There are many other observations of short-period P′ waves reflected from the 660-km discontinuity, which indicate that this discontinuity is sharp. However, the spectra of P660s are shifted to longer periods relative to P (e.g., Petersen et al., 1993) with implication that the 660–km boundary is a few tens of kilometers wide. This controversy can be reconciled with the data of mineral physics on the two phase transitions in the same depth range.

Analysis of high-quality PRFs demonstrates that the large, in a range of a few seconds, variations of travel times of the TZ phases are caused mainly by lateral heterogeneity of the upper mantle outside the TZ (Chevrot et al., 1999). If the most heterogeneous structures are excluded from consideration, these variations are similar for P660s and P410s and can be removed by taking the difference between the P660s and P410s arrival times at the same station. Lateral variations of this difference are sensitive to changes in the TZ thickness. The best data on the P660s-P410s time difference suggest that the depths of the 410- and 660-km discontinuities are stable within several kilometers everywhere, except the anomalously hot and cold regions: hotspots and subduction zones, respectively (Chevrot et al., 1999). These normal or standard depths of the major boundaries practically coincide with those in the IASP91 model (410 and 660 km, Kennett and Engdahl, 1991), and the normal TZ thickness is 250 km. This should not be mistaken for the "average" thickness.

The results for hotspots, most of which are located in oceans, are controversial (e.g., Lawrence and Shearer, 2006; Tauzin et al., 2008; Deuss, 2007; Li et al., 2003). The related data are highly variable in quality. Therefore, instead of relying on statistics, I prefer to single out the results for Iceland, which is one of the most thoroughly investigated hotspots. The network of seismograph stations in Iceland is large, and there are several independent tomographic studies of this region. This means that the topography on the 410-km discontinuity can be separated from the effects of volumetric velocity variations. Seismic tomography reveals beneath Iceland a narrow (~200 km) columnar, low-velocity body extending into the TZ (e.g., Foulger et al., 2000), in which the 410-km discontinuity is depressed by ~15 km (Shen et al., 1998), whilst the 660-km discontinuity is at its normal depth (Du et al., 2006). No TZ anomaly could be found in the data of the station BORG in Iceland (Chevrot et al., 1999), just because the columnar body is narrow, and the station is located unfavorably with respect to it. The depressed 410-km discontinuity suggests that the TZ temperature is elevated by ~150°C. The standard depth of the 660 km discontinuity implies that either the temperature at this depth is normal, or, as indicated by high-pressure experiments (Hirose, 2002), a sensitivity of depth of the transition to temperature is low in this temperature range. Note, that the accurate estimate of depth of the 660-km discontinuity in the columnar body beneath Iceland required data on the P and S velocities in this body. For most other hotspots such data are unavailable. For the ray paths outside the columnar body, the time difference between P660s and P410s is close to the normal time with implication that the TZ of the normal oceanic mantle has the same 250-km thickness as beneath the continents.

Broadly similar results are obtained for several other hotspots. For example, similar measurements at several stations in the region of the South Pacific superswell (Suetsugu et al., 2007) show that beneath one station the thickness of the TZ is 216 ± 19 km. The anomalous region cannot be larger than a few hundred kilometers. The average for all other stations is 248 ± 5 km, very close to the standard value of 250 km. By comparison, the SS-precursor data portray in the South Pacific a large (several thousand kilometers) region, where the TZ thickness is reduced to ~230 km (Flanagan and Shearer, 1998; Gu et al., 1998; Houser et al., 2008). A similar region is found in the Atlantic. These data, if accepted at face value, imply that the TZ beneath oceans differs from that beneath continents, and the TZ beneath the hotspots is the normal oceanic TZ, contrary to the receiver function data. Substantial discrepancies between the results of the two methods exist in continental regions where numerous seismograph stations facilitate good-quality receiver function studies.

Subduction zones demonstrate another kind of anomaly. Temperature in subducted slabs is anomalously low and the equilibrium depth of the olivine-wadsleyite phase transition should be ~100 km less than in ambient mantle. However, the 410-km discontinuity in most of the presently active subduction zones apparently cannot be detected with either PRFs or other methods, most likely because the olivine-wadsleyite phase transformation is kinetically hindered and a wedge-shaped zone of olivine may persist at depths greatly exceeding 410 km (Kirby

et al., 1996). The signals from the 410-km discontinuity that are sometimes obtained in seismic studies of these regions are most likely generated outside the cold subducted lithosphere. Observations of short-period phases converted from S to P near the hypocenters of deep earthquakes show that the 660-km discontinuity in subducted slabs varies in depth but sometimes is depressed by up to ~50 km (e.g., Bock and Ha, 1984; Vidale and Benz, 1992; Wicks and Richards, 1993). This can be explained by the negative Clapeyron slope of the post-spinel transition and a low temperature of the subducted slab. The temperature anomalies can be up to ~600°C. Accurate calculations are hampered by uncertainty in the estimates of the Clapeyron slope. The studies conducted with high-resolution techniques (e.g., Niu and Kawakatsu, 1998) demonstrate that the lateral scales of the anomalies are many times smaller than imaged by SS precursors.

Robust evidence for an elevated differential time P660s-P410s and a depression of the 660-km discontinuity is provided by PRFs in the Alps where the active phase of Cenozoic subduction is over, but remnants of the oceanic lithosphere are present in the TZ (Lombardi et al., 2009). Similar anomalies are either found or can be expected in other regions of the Alpine belt.

The issue of water in the TZ

The cited results of mineral physics were obtained for the dry TZ. Many researchers examined solubility of water in minerals of the TZ (for reviews see, e.g., Bolfan-Casanova, 2005; Ohtani and Sakai, 2008; Smyth and Jacobsen, 2006). Hydration means modification of structures of the TZ minerals by incorporation of hydroxyl (OH). It appears that wadsleyite and ringwoodite may incorporate up to ~2.0 wt% of water. Pressure of the olivine-wadsleyite transition decreases by hydration by up to about 1 GPa (~30 km in depth), whereas the pressure of the post-spinel transition increases. The width of the two-phase loop between olivine and wadsleyite increases with increasing water content, and may reach a few tens of kilometers in depth relative to several kilometers for dry conditions. Hydration of 1 wt% lowers S velocity by 1–2%, whereas P velocities remain practically the same.

The expected anomalies in the depth (30 km) and thickness (a few tens of kilometers) of the 410-km discontinuity are sufficiently large to be easily detected by PRFs. However seismic observations of such effects in good-quality PRFs are practically unknown. In some regions, the seismic data for the TZ are very complicated, and hydration can be one of the reasons for the complexity, but this is hard to prove. Nolet and Zielhuis (1994) reported observations of anomalously low S velocity at depths of 300–500 km beneath the Tornquist-Teisseyre zone near the western boundary of the Russian platform. They attributed this effect to hydration at the time of closure of the Paleozoic Tornquist Ocean. This observation has been made by using surface waves. The anomalous area clearly presents a good target for application of other methods, but no such study is known yet. The lack of seismic evidence for the hydrated TZ is consistent with measurements of conductivity in the TZ (Yoshino et al., 2007), which suggest that the TZ is practically dry.

A low S velocity layer atop the 410-km discontinuity may present another possible effect of hydration of the TZ. Owing to the large water solubility, the TZ may have higher water concentration than water storage capacity of the upper mantle. Then upwelling mantle material entering the upper mantle from the TZ may undergo dehydration melting (e.g., Bercovici and Karato, 2003). Huang et al. (2005) argued that even ~0.1–0.2 wt% of water in the TZ of the Pacific is sufficient for partial melting at ~410-km depth.

The layer a few tens of kilometers thick with the S velocity reduction of a few percent atop the 410-km discontinuity was detected by Revenauh and Sipkin (1994) beneath eastern China and by Vinnik et al. (1996) beneath the Kaapvaal craton in southern Africa, from multiple ScS reverberations and PRFs, respectively. Both observations were confirmed by S receiver functions for the same locations and reproduced at a number of other locations that include Antarctica, Siberia, northern Africa, and Arabian Peninsula (Vinnik and Farra, 2007). This layer is also found in the west of North America (e.g., Song et al., 2004; Jasbinsek and Dueker, 2007). Most locations of this layer seem to be associated with Cenozoic and Mesozoic mantle upwellings (Vinnik and Farra, 2007), but hydrous melting as the reason for the low S velocity still is not proved. It might be proved by observations of anomalously high anelastic attenuation, but indications of it so far were reported only for southern Africa (Vinnik et al., 2009).

Beyond the layer atop the TZ, there are indications of a low S-wave velocity layer in a depth range between 450 and 520 km at several locations (e.g., Vinnik et al., 2009). This phenomenon might also be related to hydration and it requires further analysis.

Summary

Seismic observations of the major TZ discontinuities near 410- and 660-km depths on a global scale are broadly consistent with the pyrolite mantle model. The 410-km discontinuity is related to the olivine-wadsleyite phase transition with a positive Clapeyron slope. The 660-km discontinuity is related to the post-spinel transition in the olivine component and post-majorite transition in the other components, with a negative and positive Clapeyron slope, respectively. In the normal mantle, the depths of the discontinuities are stable and in good agreement with the IASP91 model, where the thickness of the TZ is 250 km. Anomalous topography of the TZ discontinuities is related to hot and cold regions (hotspots and subduction zones, respectively). The anomalies beneath oceans are related mainly to hotspots, where the 410-km discontinuity is depressed by ~20 km. The corresponding temperature anomalies are up to ~200°C. The lack of comparable topography on the 660-km discontinuity beneath hotspots

can be explained, among other reasons, by peculiarities of phase transitions near 660-km depth. Beneath continents, except Africa and Antarctica, the anomalies are related mainly to Cenozoic subduction zones, where the 660-km discontinuity can be depressed by up to ~50 km. The related temperature anomalies are in a range of several hundred degrees Celsius. In spite of high water solubility in wadsleyite and ringwoodite, credible observations of seismic effects of hydration in the TZ are practically unknown. The thin low-S-velocity layer atop the 410-km discontinuity, found at a number of locations, is probably the only exception. The low velocity can be an effect of a hydrous melt. On a regional scale the TZ may contain less well-understood complexities such as the 520-km discontinuity or a low S velocity layer between 450- and 520-km depths.

Bibliography

Anderson, D., 1965. Recent evidence concerning the structure and composition of the Earth's mantle. *Physics and Chemistry of the Earth*, **6**, 1–131.

Benz, H. M., and Vidale, J. E., 1993. Sharpness of upper-mantle discontinuities determined from high-frequency reflections. *Nature*, **365**, 147–150, doi:10.1038/365147a0.

Bercovici, D., and Karato, S.-I., 2003. Whole-mantle convection and the transition-zone filter. *Nature*, **425**, 39–44.

Bock, G., 1994. Synthetic seismogram images of upper mantle structure: No evidence for a 520-km discontinuity. *Journal of Geophysical Research*, **99**, 15843–15851.

Bock, G., and Ha, J., 1984. Short-period S-P conversion in the mantle at a depth near 700 km. *Geophysical Journal of the Royal Astronomical Society*, **77**(2), 593–615, doi:10.1111/j.1365-246X.1984.tb01951.x.

Bolfan-Casanova, N., 2005. Water in the Earth's mantle. *Mineralogical Magazine*, **69**(3), 229–257, doi:10.1180/0026461056930248.

Cammarano, F., Deuss, A., Goes, S., and Girardini, D., 2005. One-dimensional physical reference models for the mantle and transition zone: Combined seismic and mineral physics constraints. *Journal of Geophysical Research*, **110**, B01306, doi:10.1029/2004JB003272.

Chevrot, S., Vinnik, L., and Montagner, J.-P., 1999. Global scale analysis of the mantle Pds phases. *Journal of Geophysical Research*, **104**, 20203–20219.

Deuss, A., and Woodhouse, J., 2001. Seismic observations of splitting of the mid-transition zone discontinuity in Earth's mantle. *Science*, **294**(5541), 354–357.

Deuss, A., 2007. Seismic observations of transition-zone discontinuities beneath hotspot locations. In Foulger, G. R., and Jurdy, D. M. (eds.), *Plates, Plumes, and Planetary Processes*. Geological Society of America Special paper 430, pp. 121–131.

Du, Z., Vinnik, L. P., and Foulger, G. R., 2006. Evidence from P-to-S mantle converted waves for a flat "660-km" discontinuity beneath Iceland. *Earth and Planetary Science Letters*, **241**, 271–280.

Farra, V., and Vinnik, L. P., 2000. Upper mantle stratification by P and S receiver functions. *Geophysical Journal International*, **141**, 699–712.

Flanagan, M. P., and Shearer, P. M., 1998. Global mapping of topography of transition zone velocity discontinuities by stacking SS precursors. *Journal of Geophysical Research*, **103**, 2673–2692.

Foulger, G. R., Pritchard, M. J., Julian, B. R., Evans, J. R., Allen, R. M., Nolet, G., Morgan, W. J., Bergsson, B. H., Erlendsson, P., Jacobsdottir, S., Ragnarsson, S., Stefansson, R., and Vogfjord, K., 2000. The seismic anomaly beneath Iceland extends down to the mantle transition zone and no deeper. *Geophysical Journal International*, **142**, F1–F5.

Galitzin, B., 1917. Sur l' angle d'emergence des rayons sismiques, Nachr. D. seism. Komm., 7, Heft 2, Petersburg.

Grand, S. P., and Helmberger, D., 1984. Upper mantle shear structure of North America. *Geophysical Journal of the Royal Astronomical Society*, **76**(2), 399–438.

Gu, Y. J., Dziewonski, A. M., and Agee, C. B., 1998. Global decorrelation of the topography of transition zone discontinuities. *Earth and Planetary Science Letters*, **157**, 57–67.

Hirose, K., 2002. Phase transitions in pyrolitic mantle around 670-km depth: Implications for upwelling of plumes from the lower mantle. *Journal of Geophysical Research*, **107**(B4), 2078, doi:10.1029/2001JB000597.

Houser, C., Masters, G., Flanagan, M., and Shearer, P., 2008. Determination and analysis of long-wavelength transition zone structure using SS precursors. *Geophysical Journal International*, **174**, 178–194, doi:10.1111/j.1365-246x.2008.03719.x.

Huang, X., Xu, Y., and Karato, S., 2005. Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite. *Nature*, **434**, 746–749, doi:10.1038/nature03426.

Ita, J., and Stixrude, L., 1992. Petrology, elasticity, and composition of the mantle transition zone. *Journal of Geophysical Research*, **97**, 6849–6866.

Jasbinsek, J., and Dueker, K., 2007. Ubiquitous low-velocity layer atop the 410-km discontinuity in the northern Rocky Mountains. *Geochemistry, Geophysics, Geosystems*, **8**, Q10004, doi:10.1029/2007GC001661.

Katsura, T., Yamada, H., Nishikawa, O., Song, M., Kubo, A., Shinmei, T., Yokoshi, S., Aizawa, Y., Yoshino, T., Walter, M. J., Ito, E., and Funakoshi, K., 2004. Olivine-wadsleyite transformation in the system $(Mg,Fe)_2SiO_4$. *Journal of Geophysical Research*, **109**, B02209, doi:10.1029/2003JB002438.

Kennett, B. L. N., and Engdahl, E. R., 1991. Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, **105**, 429–465.

Kirby, S. H., Stein, S., Okal, E. A., and Rubie, D. C., 1996. Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Reviews of Geophysics*, **34**(2), 261–306.

Lawrence, J. F., and Shearer, P. M., 2006. Constraining seismic velocity and density for the mantle transition zone with reflected and transmitted waveforms. *Geochemistry, Geophysics, Geosystems*, **7**, 1–19, doi:10.1029/2006GC001339.

Li, X., Kind, R., and Yuan, X., 2003. Seismic study of upper mantle and transition zone beneath hotspots. *Physics of the Earth and Planetary Interiors*, **136**, 79–92, doi:10.1016/S0031-9201(03)00021-9.

Lombardi, D., Braunmiller, J., Kissling, E., and Girardini, D., 2009. Alpine mantle transition zone imaged by receiver functions. *Earth and Planetary Science Letters*, **278**, 163–174.

Niu, F., and Kawakatsu, H., 1998. Determination of the absolute depths of the mantle transition zone discontinuities beneath China: Effect of stagnant slabs on transition zone discontinuities. *Earth Planets Space*, **50**, 965–975.

Nolet, G., and Zielhuis, A., 1994. Low S velocities under the Tornquist-Teisseyre zone: evidence for water injection into the transition zone by subduction. *Journal of Geophysical Research*, **99**, 15813–15820.

Ohtani, E., and Sakai, T., 2008. Recent advances in the study of mantle phase transitions. *Physics of the Earth and Planetary Interiors*, **170**, 240–247.

Petersen, N., Vinnik, L., Kosarev, G., Kind, R., Oreshin, S., and Stammler, K., 1993. Sharpness of the mantle discontinuities. *Geophysical Research Letters*, **20**(9), 859–862.

Revenaugh, J., and Jordan, T., 1991. Mantle layering from ScS reverberations 1. Waveform inversion of zeroth-order reverberations. *Journal of Geophysical Research*, **96**(B12), 19749–19762.
Revenaugh, J., and Sipkin, S., 1994. Seismic evidence for silicate melt atop the 410-km discontinuity. *Nature*, **369**, 474–476, doi:10.1038/369474a0.
Ritsema, J., Xu, W., Stixrude, L., and Lithgow-Bertelloni, C., 2009. Estimates of the transition zone temperature in a mechanically mixed upper mantle. *Earth and Planetary Science Letters*, **277**, 244–252.
Ryberg, T., Wenzel, F., Egorkin, A. V., and Solodilov, L., 1997. Short-period observation of the 520 km discontinuity in northern Eurasia. *Journal of Geophysical Research*, **102**(B3), 5413–5422.
Shearer, P., 1991. Constraints on upper mantle discontinuities from observations of long period reflected phases. *Journal of Geophysical Research*, **96**, 18147–18182.
Shen, Y., Solomon, S. C., Bjarnason, I Th, and Wolfe, C. J., 1998. Seismic evidence for a lower-mantle origin of the Iceland plume. *Nature*, **395**, 62–65.
Smyth, J. R., and Jacobsen, S. D., 2006. Nominally anhydrous minerals and Earth's deep water cycle. In Jacobsen, S. D., and van der Lee, S. (eds.), *Earth's Deep Water Cycle*. Washington, DC: AGU. Geophysical Monograph Series 168, doi:10.1029/168GM02.
Song, T. A., Helmberger, D. W., and Grand, S. P., 2004. Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States. *Nature*, **427**, 530–533, doi:10.1038/nature02231.
Suetsugu, D., Shiobara, H., Sigioka, H., Fukao, Y., and Kanazawa, T., 2007. Topography of the mantle discontinuities beneath the South Pacific superswell as inferred from broadband waveforms on seafloor. *Physics of the Earth and Planetary Interiors*, **160** (3–4), 310–318.
Tauzin, B., Debaule, E., and Wittlinger, G., 2008. The mantle transition zone as seen by global Pds phases: no clear evidence for a thin transition zone beneath hotspots. *Journal of Geophysical Research*, **113**, B08309, doi:10.1029/2007JB005364.
Vidale, J. E., and Benz, H. M., 1992. Upper-mantle seismic discontinuities and the thermal structure of subduction zones. *Nature*, **365**, 678–683.
Vinnik, L., 1977. Detection of waves converted from P to SV in the mantle. *Physics of the Earth and Planetary Interiors*, **15**, 39–45.
Vinnik, L., and Farra, V., 2007. Low S velocity atop the 410-km discontinuity and mantle plumes. *Earth and Planetary Science Letters*, **262**(3–4), 398–412.
Vinnik, L. P., Green, R. W. E., and Nicolaysen, L. O., 1996. Seismic constraints on dynamic of the mantle of the Kaapvaal craton. *Physics of the Earth and Planetary Interiors*, **95**, 139–151.
Vinnik, L., Oreshin, S., Kosarev, G., Kiselev, S., and Makeyeva, L., 2009. Mantle anomalies beneath southern Africa: Evidence from seismic S and P receiver functions. *Geophysical Journal International*, **179**, 279–298.
Wicks, W., and Richards, M. A., 1993. A detailed map of the 660-km discontinuity beneath the Izu-Bonin subduction zone. *Science*, **261**(5127), 1424–1427, doi:10.1126/science.261.5127.1424.
Yoshino, T., Manthilake, G., Matsuzaki, T., and Katsura, T., 2007. Dry mantle transition zone inferred from the conductivity of wadsleyite and ringwoodite. *Nature*, **451**, 326–329, doi:10.1038/nature06427.

Cross-references

Body Waves
Earth's Structure, Global
Earth's Structure, Upper Mantle
Geodynamics
Mantle Convection
Mantle Plumes
Seismic Phase Names: IASPEI Standard
Seismic, Receiver Function Technique
Seismology, Global Earthquake Model
Subduction Zones

SEISMIC HAZARD

Andrzej Kijko
Aon-Benfield Natural Hazard Centre, University of Pretoria, Pretoria, Republic of South Africa

Definition

Seismic hazard. Any physical phenomena associated with an earthquake (e.g., ground motion, ground failure, liquefaction, and tsunami) and their effects on land, man-made structure, and socioeconomic systems that have the potential to produce a loss. It is also used without regard to a loss to indicate the probable level of ground shaking occurring at a given point within a certain period of time.
Seismic hazard analysis. Quantification of the ground motion expected at a particular site.
Deterministic seismic hazard analysis. Quantification of a single or relatively small number of individual earthquake scenarios.
Probabilistic seismic hazard analysis. Quantification of the probability that a specified level of ground motion will be exceeded at least once at a site or in a region during a specified exposure time.
Ground motion prediction equation. A mathematical equation which indicates the relative decline of the ground motion parameter as the distance from the earthquake increases.

Introduction

The estimation of the expected ground motion which can occur at a particular site is vital to the design of important structures such as nuclear power plants, bridges, and dams. The process of evaluating the design parameters of earthquake ground motion is called seismic hazard assessment or seismic hazard analysis. Seismologists and earthquake engineers distinguish between seismic hazard and seismic risk assessments in spite of the fact that in everyday usage these two phrases have the same meaning. Seismic hazard is used to characterize the severity of ground motion at a site regardless of the consequences, while the risk refers exclusively to the consequences to human life and property loss resulting from the occurred hazard. Thus, even a strong earthquake can have little risk potential if it is far from human development and infrastructure, while a small seismic event in an unfortunate location may cause extensive damage and losses.

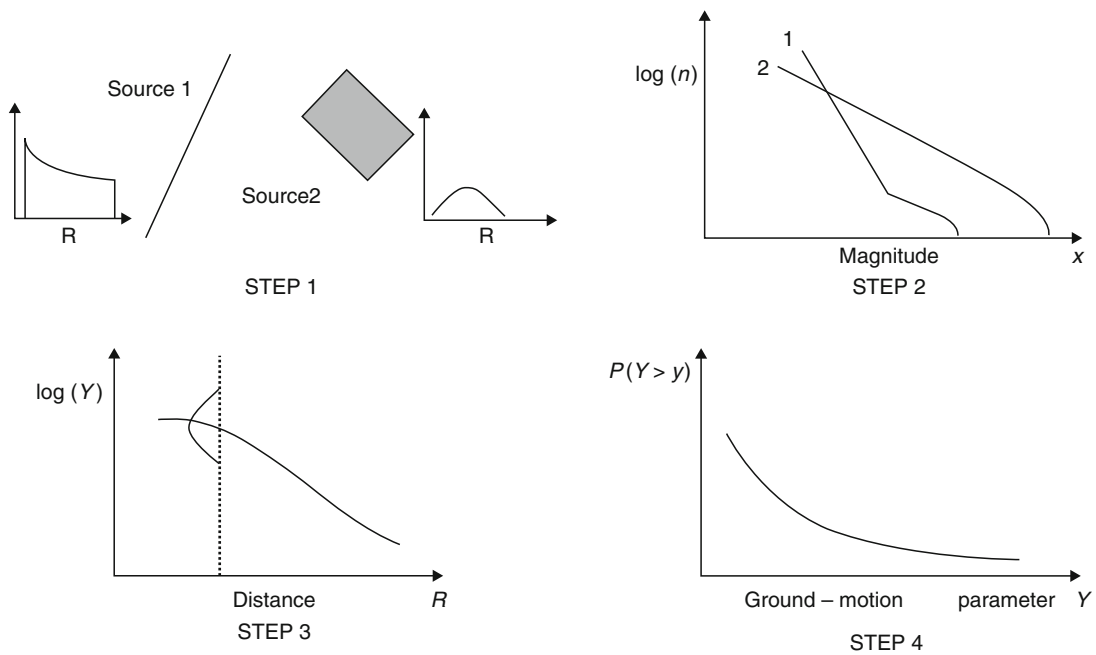
Seismic hazard analysis can be performed *deterministically*, when a particular earthquake scenario is considered, or *probabilistically*, when likelihood or frequency of specified earthquake size and location are evaluated.

The process of *deterministic* seismic hazard analysis (DSHA) involves the initial assessment of the maximum possible earthquake magnitude for each of the various seismic sources such as active faults or seismic source zones (SSHAC, 1997). An area of up to 450 km radius around the site of interest can be investigated. Assuming that each of these earthquakes will occur at the minimum possible distance from the site, the ground motion is calculated using appropriate attenuation equations. Unfortunately, this straightforward and intuitive procedure is overshadowed by the complexity and uncertainty in selecting the appropriate earthquake scenario, creating the need for an alternative, *probabilistic* methodology, which is free from discrete selection of scenario earthquakes. Probabilistic seismic hazard analysis (PSHA) quantifies as a probability whatever hazard may result from all earthquakes of all possible magnitudes and at all significant distances from the site of interest. It does this by taking into account their frequency of occurrence. Deterministic earthquake scenarios, therefore, are a special case of the probabilistic approach. Depending on the scope of the project, DSHA and PSHA can complement one another to provide additional insights to the seismic hazard (McGuire, 2004). This study will concentrate on a discussion of PSHA.

In principle, any natural hazard caused by seismic activity can be described and quantified by the formalism of the PSHA. Since the damages caused by ground shaking very often result in the largest economic losses, our presentation of the basic concepts of PSHA is illustrated by the quantification of the likelihood of ground shaking generated by earthquakes. Modification of the presented formalism to quantify any other natural hazard is straightforward.

The classic procedure for the PSHA includes four steps (Reiter, 1990), (Figure 1).

1. The first step consists of the identification and parameterization of the *seismic sources* (known also as *source zones*, *earthquake sources*, or *seismic zones*) that may affect the site of interest. These may be represented as area, fault, or point sources. Area sources are often used when one cannot identify a specific fault. In classic PSHA, a uniform distribution of seismicity is assigned to each earthquake source, implying that earthquakes are equally likely to occur at any point within the source zone. The combination of earthquake occurrence distributions with the source geometry, results in space, time, and magnitude distributions of earthquake occurrences. Seismic source models can be interpreted as a list of potential scenarios, each with an associated magnitude, location, and seismic activity rate (Field, 1995).
2. The next step consists of the specification of temporal and magnitude distributions of seismicity for each source. The classic Cornell–McGuire approach assumes that earthquake occurrence in time is random and follows the Poisson process. This implies that earthquake occurrences in time are statistically independent and that they occur at a constant rate. Statistical independence means that occurrence of future earthquakes does not depend on the occurrence of the past earthquake. The most often used model of earthquake magnitude recurrence is the



Seismic Hazard, Figure 1 Four steps of a PSHA (Modified from Reiter, 1990).

frequency-magnitude Gutenberg–Richter relationship (Gutenberg and Richter, 1944).

$$\log(n) = a - bm, \tag{1}$$

where n is the number of earthquakes with a magnitude of m and a and b are parameters. It is assumed that earthquake magnitude, m, belongs to the domain $<m_{\min}, m_{\max}>$, where $m_{\min}$ is the level of completeness of earthquake catalogue and magnitude $m_{\max}$ is the upper limit of earthquake magnitude for a given seismic source. The parameter a is the measure of the level of seismicity, while b describes the ratio between the number of small and large events. The Gutenberg–Richter relationship may be interpreted either as being a cumulative relationship, if n is the number of events with magnitude equal or larger than m, or as being a density law, stating that n is the number of earthquakes in a specific, small magnitude interval around m. Under the above assumptions, the seismicity of each seismic source is described by four parameters: the (annual) rate of seismicity λ, which is equal to the parameter of the Poisson distribution, the lower and upper limits of earthquake magnitude, $m_{\min}$ and $m_{\max}$, and the b-value of the Gutenberg–Richter relationship.

3. Calculation of ground motion prediction equations and their uncertainty. Ground motion prediction equations are used to predict ground motion at the site itself. The parameters of interest include peak ground acceleration, peak ground velocity, peak ground displacement, spectral acceleration, intensity, strong ground motion duration, etc. Most ground motion prediction equations available today are empirical and depend on the earthquake magnitude, source-to-site distance, type of faulting, and local site conditions. The choice of an appropriate ground motion prediction equation is crucial since, very often, it is a major contributor to uncertainty in the estimated PSHA.
4. Integration of uncertainties in earthquake location, earthquake magnitude and ground motion prediction equation into probability that the ground motion parameter of interest will be exceeded at the specified site during the specified time interval. The ultimate result of a PSHA is a *seismic hazard curve*: the annual probability of exceeding a specified ground motion parameter at least once. An alternative definition of the hazard curve is the frequency of exceedance versus ground motion amplitude (McGuire, 2004).

The following section provides the mathematical framework of the classic PSHA procedure, including its deaggregation. The most common modifications of the procedure will be discussed in the section (Some modifications of Cornell–McGuire PSHA procedure and alternative models).

The Cornell–McGuire PSHA methodology

Conceptually, the computation of a seismic hazard curve is fairly simple. Let us assume that seismic hazard is characterized by ground motion parameter Y. The probability of exceeding a specified value y, $P[Y \geq y]$, is calculated for an earthquake of particular magnitude located at a possible source, and then multiplied by the probability that that particular earthquake will occur. The computations are repeated and summed for the whole range of possible magnitudes and earthquake locations. The resulting probability $P[Y \geq y]$ is calculated by utilizing the Total Probability Theorem which is:

$$P[Y \geq y] = \sum P[Y \geq y|E_i] \cdot P[E_i], \tag{2}$$

where

$$P[Y \geq y|E_i] = \int \cdots \int P[Y \geq y|x_1, x_2, x_3 \ldots] \cdot f_i(x_1) \cdot f_i(x_2|x_1) \cdot f_i(x_3|x_1, x_2) \ \ldots \ dx_3 \, dx_2 \, dx_1. \tag{3}$$

$P[Y \geq y|E_i]$ denotes the probability of ground motion parameter $Y \geq y$, at the site of interest, when an earthquake occurs within the seismic source i. Variables x_i $(i = 1, 2, \ldots)$ are uncertainty parameters that influence Y. In the classic approach, as developed by Cornell (1968), and later extended to accommodate ground motion uncertainty (Cornell, 1971), the parameters of ground motion are earthquake magnitude, M, and earthquake distance, R. Functions $f(\cdot)$ are probability density functions (PDF) of parameters x_i. Assuming that indeed $x_1 \equiv M$ and $x_2 \equiv R$, the probability of exceedance (Equation 3) takes the form:

$$P[Y \geq y|E] = \int_{m_{\min}}^{m_{\max}} \int_{R|M} P[Y \geq y|m, r] \, f_M(m) f_{R|M}(r|m) \, dr \, dm, \tag{4}$$

where $P[Y \geq y|m, r]$ denotes the conditional probability that the chosen ground motion level y is exceeded for a given magnitude and distance; $f_M(m)$ is the PDF of earthquake magnitude, and $f_{R|M}(r|m)$ is the conditional PDF of the distance from the earthquake for a given magnitude. The $f_{R|M}(r|m)$ arises in specific instances, such as those where a seismic source is represented by a fault rupture. Since the earthquake magnitude depends on the length of fault rupture, the distance to the rupture and resulting magnitude are correlated.

If, in the vicinity of the site of interest, one can distinguish n_S seismic sources, each with average annual rate of earthquake occurrence λ_i, then the total average annual rate of events with a site ground motion level y or more, takes the form:

$$\lambda(y) = \sum_{i=1}^{n_S} \lambda \int_{m_{\min}}^{m_{\max}} \int_{R|M} P[Y \geq y|M, R] \, f_M(m) f_{R|M}(r|m) \, dr \, dm, \tag{5}$$

In Equation 5, the subscripts denoting seismic source number are deleted for simplicity, $P[Y \geq y|m,r]$ denotes the conditional probability that the chosen ground motion level y, is exceeded for a given magnitude m and distance r. The standard choice for the probability $P[Y \geq y|m,r]$ is a normal, complementary cumulative distribution function (CDF), which is based on the assumption that the ground motion parameter y is a log-normal random variable, $\ln(y) = g(m,r) + \varepsilon$, where ε is random error. The mean value of $\ln(y)$ and its standard deviation are known and are defined as $\overline{\ln(y)}$ and $\sigma_{\ln(y)}$, respectively. The function $f_M(m)$ denotes the PDF of earthquake magnitude. In most engineering applications of PSHA, it is assumed that earthquake magnitudes follow the Gutenberg–Richter relation (1), which implies that $f_M(m)$ is a negative, exponential distribution, shifted from zero to $m_{\min}$ and truncated from the top by $m_{\max}$

$$f_M(m) = \frac{\beta \exp[(-(m - m_{\min})]}{1 - \exp[(-\beta(m_{\max} - m_{\min})]}, \quad (6)$$

In Equation 6, $\beta = b \ln 10$, where b is the parameter of the frequency-magnitude Gutenberg–Richter relation (1).

After assuming that in every seismic source, earthquake occurrences in time follow a Poissonian distribution, the probability that y, a specified level of ground motion at a given site, will be exceeded at least once within any time interval t is

$$P[Y > y;\ t] = 1 - \exp[-\lambda(y) \cdot t]. \quad (7)$$

The Equation 7 is fundamental to PSHA. For $t = 1$ year, its plot versus ground motion parameter y, is the *hazard curve* – the ultimate product of the PSHA (Figure 2). For small probabilities,

$$P[Y > y; t = 1] = 1 - \exp(-\lambda)$$
$$\cong 1 - (1 - \lambda + \frac{1}{2}\lambda^2 - ...) \cong \lambda, \quad (8)$$

which means that the probability (Equation 7) is approximately equal to $\lambda(y)$. This proves that PSHA can be characterized interchangeably by the annual probability (Equation 7) or by the rate of seismicity (Equation 5).

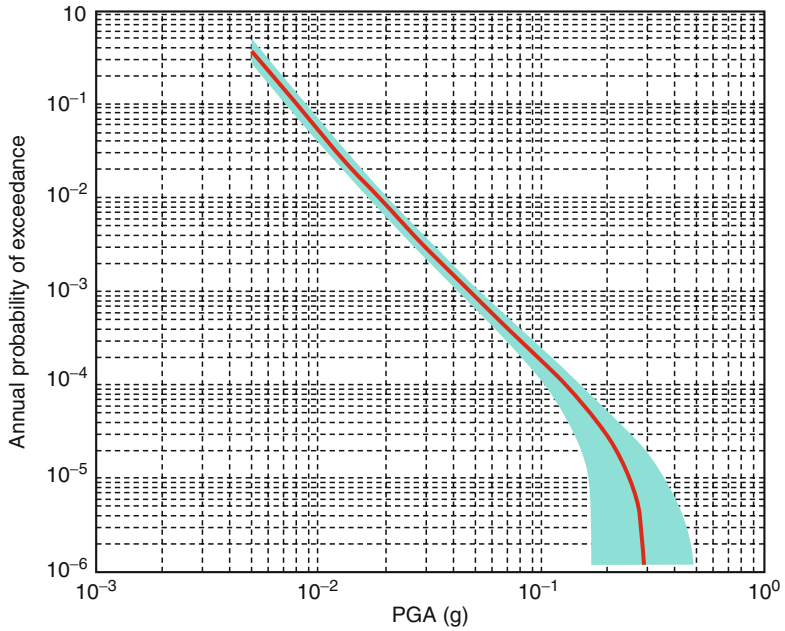
In the classic Cornell–McGuire procedure for PSHA, it is assumed that the earthquakes in the catalogue are independent events. The presence of clusters of seismicity, multiple events occurring in a short period of time or presence of foreshocks and aftershocks violates this assumption. Therefore, before computation of PSHA, these dependent events must be removed from the catalogue.

Estimation of seismic source parameters

Following the classic Cornell–McGuire PSHA procedure, each seismic source is characterized by four parameters:

- Level of completeness of the seismic data, $m_{\min}$
- Annual rate of seismic activity λ, corresponding to magnitude $m_{\min}$
- b-value of the frequency-magnitude Gutenberg–Richter relation (1)
- Upper limit of earthquake magnitude $m_{\max}$

Estimation of m_{min}. The level of completeness of the seismic event catalogue, $m_{\min}$, can be estimated in at least two different ways.



Seismic Hazard, Figure 2 Example of a Peak Ground Acceleration (PGA) seismic hazard curve and its confidence intervals.

The first approach is based on information provided by the seismic event catalogue itself, where $m_{\min}$ is defined as the deviation point from an empirical or assumed earthquake magnitude distribution model. Despite the fact that the evaluation of $m_{\min}$ based on information provided entirely by seismic event catalogue is widely used, it has several weak points. By definition, the estimated levels of $m_{\min}$ represent only the average values over space and time. However, most procedures in this category require assumptions on a model of earthquake occurrence, such as a Poissonian distribution in time and frequency-magnitude Gutenberg–Richter relation.

The second approach used for the estimation of $m_{\min}$ utilizes information on the detection capabilities of seismic stations. The approach release users from the assumptions of stationarity and statistical independence of event occurrence. The choice of the most appropriate procedure for $m_{\min}$ estimation depends on several factors, such as the knowledge of the history of the development of the seismic network, data collection, and processing.

Estimation of rate of seismic activity λ and b-value of Gutenberg-Richter. The accepted approach to estimating seismic source recurrence parameters λ and b is the maximum likelihood procedure. If successive earthquakes are independent in time, the number of earthquakes with magnitude equal to or exceeding a level of completeness, $m_{\min}$, follows the Poisson distribution with the parameter equal to the annual rate of seismic activity λ. The maximum likelihood estimator of λ is then equal to n/t, where n is number of events that occurred within time interval t.

For given $m_{\max}$, the maximum likelihood estimator of the b-value of the Gutenberg–Richter equation can be obtained from the recursive solution of the following:

$$1/\beta = \bar{m} - m_{\min} + \frac{(m_{\max} - m_{\min}) \cdot \exp[-\beta(m_{\max} - m_{\min})]}{1 - \exp[-\beta(m_{\max} - m_{\min})]}. \quad (9)$$

where $\beta = b \ln 10$, and $\bar{m}$ is the sample mean of earthquake magnitude. If the range of earthquake magnitudes $< m_{\max}, m_{\min} >$ exceeds 2 magnitude units, the solution of Equation 9 can be approximated by the well-known Aki-Utsu estimator (Aki, 1965; Utsu, 1965)

$$\beta = 1 \,/\, (\bar{m} - m_{\min}). \quad (10)$$

In most real cases, estimation of parameters λ and the b-value by the above simple formulas cannot be performed due to the incompleteness of seismic event catalogues. The alternative procedures, capable to utilize data incompleteness has been developed by Weichert (1980) and Kijko and Sellevoll (1992).

Estimation of m_{max}. The maximum magnitude, $m_{\max}$, is defined as the upper limit of magnitude for a given seismic source.

This terminology assumes a sharp cutoff magnitude at a maximum magnitude $m_{\max}$. Cognizance should be taken of the fact that an alternative, "soft" cutoff maximum earthquake magnitude is also being used (Main and Burton, 1984). The later formalism is based on the assumption that seismic moments of seismic events follow the Gamma distribution. One of the distribution parameters is called the maximum seismic moment, and the corresponding value of earthquake magnitude is called the "soft" maximum magnitude. Beyond the value of this maximum magnitude, the distribution decays much faster than the classical Gutenberg–Richter relation. Although this model has been occasionally used, the classic PSHA only considers models having a sharp cutoff of earthquake magnitude.

As a rule, $m_{\max}$ plays an important role in PSHA, especially in assessment of long return periods. At present, there is no generally accepted method for estimating $m_{\max}$. It is estimated by the combination of several factors, which are based on two kinds of information: seismicity of the area, and geological, geophysical, and structural information of the seismic source. The utilization of the seismological information focuses on the maximum observed earthquake magnitude within a seismic source and statistical analysis of the available seismic event catalogue. The geological information is used to identify distinctive tectonic features, which control the value of $m_{\max}$.

The current evaluations of $m_{\max}$ are divided between deterministic and probabilistic procedures, based on the nature of the tools applied.

Deterministic procedures. The deterministic procedure most often applied is based on the empirical relationships between magnitude and various tectonic and fault parameters, such as fault length or rupture dimension. The relationships are different for different seismic areas and different types of faults (Wells and Coppersmith, 1994 and references therein). Despite the fact that such empirical relationships are extensively used in PSHA (especially for the assessment of maximum possible magnitude generated by the fault-type seismic sources), the weak point of the approach is its requirement to specify the highly uncertain length of the future rupture. An alternative approach to the determination of earthquake recurrence on singular faults with a segment specific slip rate is provided by the so-called cascade model, where segment rupture is defined by the individual cascade-characteristic rupture dimension (Cramer et al., 2000).

Another deterministic procedure which has a strong, intuitive appeal is based on records of the largest historic or paleo-earthquakes (McCalpin, 1996). This approach is especially applicable in the areas of low seismicity, where large events have long return periods. In the absence of any additional tectono-geological indications, it is assumed that the maximum possible earthquake magnitude is equal to the largest magnitude observed, $m_{\max}^{obs}$, or the largest observed plus an increment. Typically, the increment varies from ¼ to 1 magnitude unit. The procedure is often used for the areas with several, small seismic sources, each having its own $m_{\max}^{obs}$ (Wheeler, 2009).

Another commonly used deterministic procedure for $m_{\max}$ evaluation, especially for area-type seismic sources, is based on the extrapolation of the frequency-magnitude

Gutenberg–Richter relation. The best known extrapolation procedures are probably those by Frohlich (1998) and the "probabilistic" extrapolation procedure applied by Nuttli (1981), in which the frequency-magnitude curve is truncated at the specified value of annual probability of exceedance (e.g., 0.001).

An alternative procedure for the estimation of $m_{\max}$ was developed by Jin and Aki (1988), where a remarkably linear relationship was established between the logarithm of coda Q_0 and the largest observed magnitude for earthquakes in China. The authors postulate that if the largest magnitude observed during the last 400 years is the maximum possible magnitude $m_{\max}$, the established relation will give a spatial mapping of $m_{\max}$.

Ward (1997) developed a procedure for the estimation of $m_{\max}$ by simulation of the earthquake rupture process. Ward's computer simulations are impressive; nevertheless, one must realize that all the quantitative assessments are based on the particular rupture model, postulated parameters of the strength and assumed configuration of the faults.

The value of $m_{\max}$ can also be estimated from the tectono-geological features like strain rate or the rate of seismic-moment release (WGCEP, 1995). Similar approaches have also been applied in evaluating the maximum possible magnitude of seismic events induced by mining (e.g., McGarr, 1984). However, in most cases, the uncertainty of $m_{\max}$ as determined by any deterministic procedure is large, often reaching a value of the order of one unit on the Richter scale.

Probabilistic procedures. The first probabilistic procedure for maximum regional earthquake magnitude was developed in the late 1960s, and is based on the formalism of the extreme values of random variables. A major breakthrough in the seismological applications of extreme-value statistics was made by Epstein and Lomnitz (1966), who proved that the Gumbel I distribution of extremes can be derived directly from the assumptions that seismic events are generated by a Poisson process and that they follow the frequency-magnitude Gutenberg–Richter relation. Statistical tools required for the estimation of the end-point of distribution functions (as, e.g., Cooke, 1979) have only recently been used in the estimation of maximum earthquake magnitude (Pisarenko et al., 1996; Kijko, 2004 and references therein).

The statistical tools available for the estimation of $m_{\max}$ vary significantly. The selection of the most suitable procedure depends on the assumptions of the statistical distribution model and/or the information available on past seismicity. Some of the procedures can be applied in the extreme cases when no information about the nature of the earthquake magnitude distribution is available. Some of the procedures can also be used when the earthquake catalogue is incomplete, i.e., when only a limited number of the largest magnitudes are known. Two estimators are presented here. Broadly speaking, the first estimator is straightforward and simple in application, while the second one requires more computational effort but provides more accurate results. It is assumed that both the analytical form and the parameters of the distribution functions of earthquake magnitude are known. This knowledge can be very approximate, but must be available.

Based on the distribution of the largest among n observations and on the condition that the largest observed magnitude $m_{\max}^{obs}$ is equal to the largest magnitude to be expected, the "simple" estimate of $m_{\max}$ is of the form

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{1}{n f_M(m_{\max}^{obs})}, \tag{11}$$

where $f_M(m_{\max}^{obs})$ is PDF of the earthquake magnitude distribution. If applied to the Gutenberg–Richter recurrence relation with PDF (Equation 6), it takes the simple form

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{1 - \exp[-\beta(m_{\max}^{obs} - m_{\min})]}{n\beta \exp[-\beta(m_{\max}^{obs} - m_{\min})]}. \tag{12}$$

The approximate variance of the estimator (Equation 12) is of the form

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \frac{1}{n^2} \left[\frac{1 - \exp[-\beta(m_{\max}^{obs} - m_{\min})]}{\beta \exp[-\beta(m_{\max}^{obs} - m_{\min})]}\right]^2, \tag{13}$$

where σ_M stands for epistemic uncertainty and denotes the standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$. The second part of the variance represents the aleatory uncertainty of $m_{\max}$.

The second ("advanced") procedure often used for assessment of $m_{\max}$ is based on the formalism derived by Cooke (1979)

$$\hat{m}_{\max} = m_{\max}^{obs} + \int_{m_{\min}}^{m_{\max}^{obs}} [F_M(m)]^n \mathrm{d}m, \tag{14}$$

where $F_M(m)$ denotes the CDF of random variable m. If applied to the frequency-magnitude Gutenberg–Richter relation (1), the respective CDF is

$$F_M(m) = \begin{cases} 0, & \text{for } m < m_{\min}, \\ \frac{1-\exp[-\beta(m-m_{\min})]}{1-\exp[-\beta(m_{\max}-m_{\min})]}, & \text{for } m_{min} \le m \le m_{\max}, \\ 1, & \text{for } m > m_{\max}, \end{cases} \tag{15}$$

and the $m_{\max}$ estimator (Equation 14) takes the form

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{E_1(n_2) - E_1(n_1)}{\beta \exp(-n_2)} + m_{\min} \exp(-n), \tag{16}$$

where $n_1 = n/\{1 - \exp[-\beta(m_{\max}^{obs} - m_{\min})]\}$, $n_2 = n_1 \exp[-\beta(m_{\max}^{obs} - m_{\min})]$, and $E_1(\cdot)$ denotes an exponential integral function. The variance of estimator (Equation 16)

has two components, epistemic and aleatory, and is of the form

$$VAR(\hat{m}_{\max}) = \sigma_M^2 + \left[\frac{E_1(n_2) - E_1(n_1)}{\beta \exp(-n_2)} + m_{\min} \exp(-n)\right]^2, \quad (17)$$

where σ_M denotes standard error in the determination of the largest observed magnitude $m_{\max}^{obs}$.

Both above estimators of $m_{\max}$, by their nature, are very general and have several attractive properties. They are applicable for a very broad range of magnitude distributions. They may also be used when the exact number of earthquakes, n, is not known. In this case, the number of earthquakes can be replaced by λt. Such a replacement is equivalent to the assumption that the number of earthquakes occurring in unit time conforms to a Poisson distribution with parameter λ, where t is the span of the seismic event catalogue. It is also important to note that both estimators provide a value of $\hat{m}_{\max}$, which is never less than the largest magnitude already observed.

Alternative procedures are discussed by Kijko (2004), which are appropriate for the case when the empirical magnitude distribution deviates from the Gutenberg-Richter relation. These procedures assume no specific form of the magnitude distribution or that only a few of the largest magnitudes are known.

Despite the fact that statistical procedures based on the mathematical formalism of extreme values provide powerful tools for the evaluation of $m_{\max}$, they have one weak point: often available seismic event catalogues are too short and insufficient to provide reliable estimations of $m_{\max}$. Therefore the Bayesian extension of statistical procedures (Cornell, 1994), allowing the inclusion of alternative and independent information such as local geological conditions, tectonic environment, geophysical data, paleoseismicity, similarity with another seismic area, etc., are able to provide more reliable assessments of $m_{\max}$.

Numerical computation of PSHA

With the exception of a few special cases (Bender, 1984), the hazard curve (Equation 7) cannot be computed analytically. For the most realistic distributions, the integrations can only be evaluated numerically. The common practice is to divide the possible ranges of magnitude and distance into n_M and n_R intervals, respectively. The average annual rate (Equation 4) is then estimated as

$$\lambda(Y > y) \cong \sum_{i=1}^{n_S} \sum_{j=1}^{n_M} \sum_{k=1}^{n_R} \lambda_i P[Y > y | m_j, r_k] f_{M_j}(m_j) f_{R_k}(r_k) \Delta m \Delta r, \quad (18)$$

where $m_j = m_{\min} + (j - 0.5) \cdot (m_{\max} - m_{\min})/n_M$, $r_k = r_{\min} + (k - 0.5) \cdot (r_{\max} - r_{\min})/n_R$, $\Delta m = (m_{\max} - m_{\min})/n_M$, and $\Delta r = (r_{\max} - r_{\min})/n_R$.

If the procedure is applied to a grid of points, it will result in a map of PSHA, in which the contours of the expected ground motion parameter during the specified time interval can be drawn (Figure 3).

Deaggregation of seismic hazard

By definition, the PSHA aggregates ground motion contributions from earthquake magnitudes and distances of significance to a site of engineering interest. One has to note that the PSHA results are not representative of a single earthquake. However, an integral part of the design procedure of any critical structure is the analysis of the most relevant earthquake acceleration time series, which are generated by earthquakes, at specific magnitudes and distances. Such earthquakes are called "controlling earthquakes," and they are used to determine the shapes of the response spectral acceleration or PGA at the site.

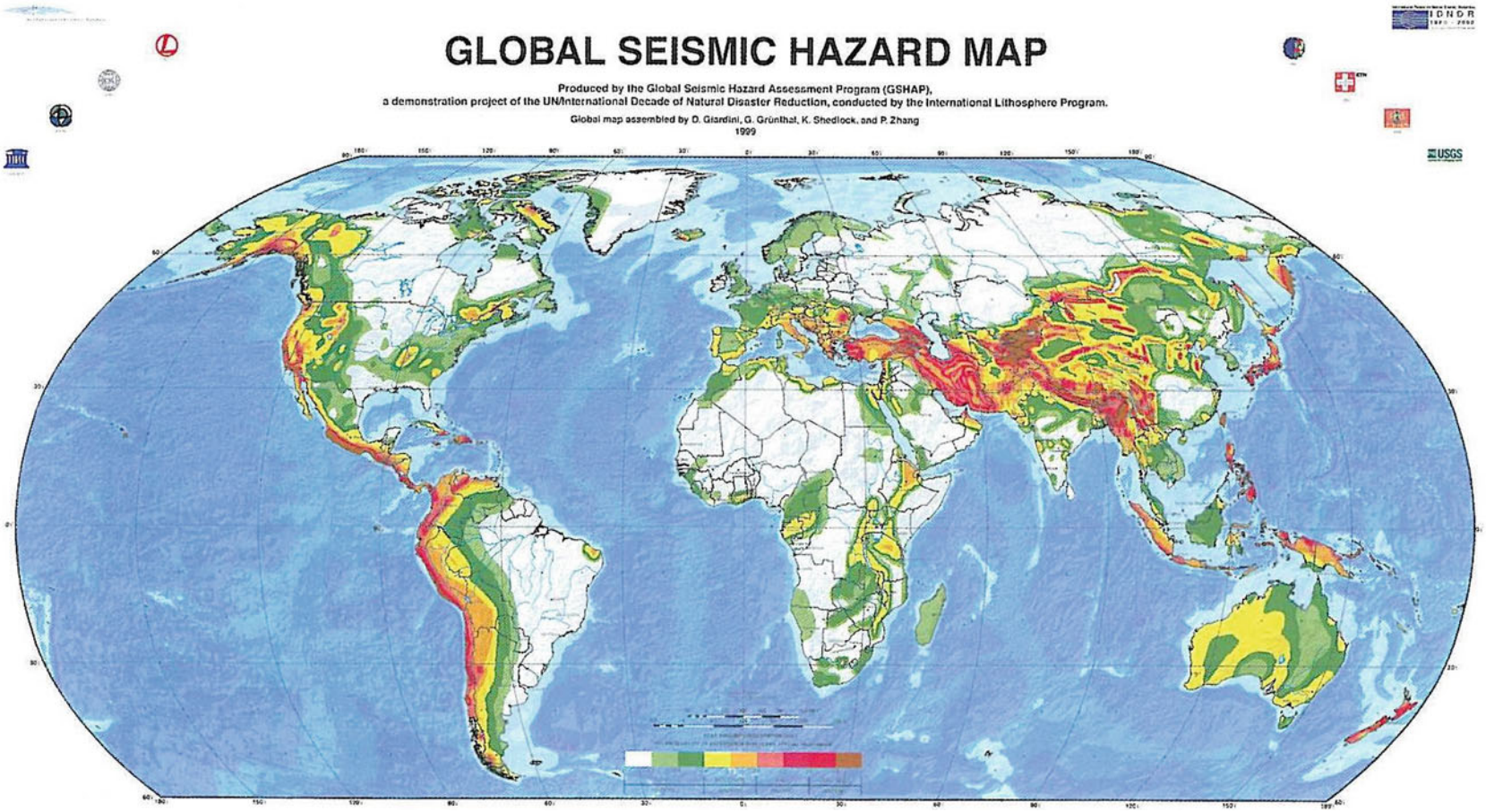
Controlling earthquakes are characterized by mean magnitudes and distances derived from so-called deaggregation analysis. During the deaggregation procedure, the results of PSHA are separated to determine the dominant magnitudes and the distances that contribute to the hazard curve at a specified (reference) probability. Controlling earthquakes are calculated for different structural frequency vibrations, typically for the fundamental frequency of a structure. In the process of deaggregation, the hazard for a reference probability of exceedance of specified ground motion is portioned into magnitude and distance bins. The relative contribution to the hazard for each bin is calculated. The bins with the largest relative contribution identify those earthquakes that contribute the most to the total seismic hazard.

Some modifications of Cornell–McGuire PSHA procedure and alternative models

Source-free PSHA procedures

The concept of seismic sources is the core element of the Cornell–McGuire PSHA procedure. Unfortunately, seismic sources or specific faults can often not be identified and mapped and the causes of seismicity are not understood. In these cases, the delineation of seismic sources is highly subjective and is a matter of expert opinion. In addition, often, seismicity within the seismic sources is not distributed uniformly, as it is required by the classic Cornell–McGuire procedure. The difficulties experienced in dealing with seismic sources have stimulated the development of an alternative technique to PSHA, which is free from delineation of seismic sources.

One of the first attempts to develop an alternative to the Cornell–McGuire procedure was made by Veneziano et al. (1984). Indeed, the procedure does not require the specification of seismic sources, is non-parametric, and, as input, requires only information about past seismicity. The empirical distribution of the specified seismic hazard parameter is calculated by using the observed earthquake magnitudes, epicentral distances, and assumed ground motion prediction equation. By normalizing this



Seismic Hazard, Figure 3 Example of product of PSHA. Map of seismic hazard of the world. Peak ground acceleration expected at 10% probability of exceedance at least once in 50 years. (From Giardini, 1999, http://www.gfz-potsdam.de/pb5/pb53/projects/gshap).

distribution for the duration of the seismic event catalogue, one obtains an annual rate of the exceedance for the required hazard parameter.

Another non-parametric PSHA procedure has been developed by Woo (1996). The procedure is also source-free, where seismicity distributions are approximated by data-based kernel functions. Comparison of the classic Cornell–McGuire-based and kernel-based procedures shows that the former yields a lower hazard.

By their nature, the non-parametric procedures work well in areas with a frequent occurrence of strong seismic events and where the record of past seismicity is considerably complete. At the same time, the non-parametric approach has significant weak points. Its primary disadvantage is a poor reliability in estimating small probabilities for areas of low seismicity. The procedure is not recommended for an area where the seismic event catalogues are highly incomplete. In addition, in its present form, the procedure is not capable of making use of any additional geophysical or geological information to supplement the pure seismological data. Therefore, a technique that accommodates the incompleteness of the seismic event catalogues and, at the same time, does not require the specification of seismic sources, would be an ideal tool for analyzing and assessing seismic hazard.

Such a technique, which can be classified as a *parametric-historic* procedure for PSHA has been successfully used in several parts of the world. Kijko (2008) used it for mapping seismic hazard of South Africa and sub-Saharan Africa. The procedure has been applied in selected parts of the world by the Global Seismic Hazard Assessment Program (GSHAP, Giardini, 1999), while Petersen et al. (2008) applied it for mapping the seismic hazard in the USA. In a series of papers, Frankel and his colleagues modified and substantially extended the original procedure. Their final approach is parametric and based on the assumption that earthquakes within a specified grid size are Poissonian in time, and that the earthquake magnitudes follow the Gutenberg–Richter relation truncated from the top by maximum possible earthquake magnitude $m_{\max}$.

In some cases, the frequency-magnitude Gutenberg–Richter relation is extended by characteristic events. The procedure accepts the contribution of seismicity from active faults and compensates for incompleteness of seismic event catalogues. Frankel's conceptually simple and intuitive parametric-historic approach combines the best of the deductive and non-parametric-historic procedures and, in many cases, is free from the disadvantages characteristic of each of the procedures. The rigorous mathematical foundations of the parametric-historic PSHA procedure have been given by Kijko and Graham (1999).

Alternative earthquake recurrence models

Time-dependent models. In addition to the classic assumption, that earthquake occurrence in time follows a Poisson process, alternative approaches are occasionally used. These procedures attempt to assess temporal, or temporal and spatial dependence of seismicity. Time-dependent earthquake occurrence models specify a distribution of the time to the next earthquake, where this distribution depends on the magnitude of the most recent earthquake. In order to incorporate the memory of past events, the non-Poissonian distributions or Markov chains are applied. In this approach, the seismogenic zones that recently produced strong earthquakes become less hazardous than those that did not rupture in recent history.

Clearly such models may result in a more realistic PSHA, but most of them are still only research tools and have not yet reached the level of development required by routine engineering applications.

Time-dependent occurrence of large earthquakes on segments of active faults is extensively discussed by Rhoades et al. (1994) and Ogata (1999) Also, a comprehensive review of all aspects of non-Poissonian models is provided by Kramer (1996). There are several time-dependent models which play an important role in PSHA. The best known models, which have both firm physical and empirical bases, are probably the models by Shimazaki and Nakata (1980). Based on the correlation of seismic activity with earthquake-related coastal uplift in Japan, Shimazaki and Nakata (1980) proposed two models of earthquake occurrence: a *time-predictable* and a *slip-predictable*.

The time-predictable model states that earthquakes occur when accumulated stress on a fault reaches a critical level; however, the stress drop and magnitudes of the subsequent earthquakes vary among seismic cycles. Thus, assuming a constant fault-slip rate, the time to the next earthquake can be estimated from the slip of the previous earthquake. The second, the slip-predictable model, is based on the assumption that, irrespective of the initial stress on the fault, an earthquake occurrence always causes a reduction in stress to the same level. Thus, the fault-slip in the next earthquake can be estimated from the time since the previous earthquake.

The second group of time-dependent models are less tightly based on the physical considerations of earthquake occurrence and attempt to describe intervals between the consecutive events by specified statistical distributions. Ogata (1999) considers five models: log-normal, gamma, Weibull, doubly exponential and exponential, which results in the stationary Poisson process. After application of these models to several paleo-earthquake data sets, he concluded that no one of the distributions is consistently the best fit; the quality of the fit strongly depends on the data. From several attempts to describe earthquake time intervals between consecutive events using statistical distributions, at least two play a significant role in the current practice of PSHA: the log-normal and the Brownian passage time (BPT) renewal model.

The use of a log-normal model is justified by the discovery that normalized intervals between the consecutive large earthquakes in the circum-Pacific region follow a log-normal distribution with an almost constant standard

deviation (Nishenko and Buland, 1987). The finite value for the intrinsic standard deviation is important because it controls the degree of aperiodicity in the occurrence of *characteristic earthquakes,* making accurate earthquake prediction impossible. Since this discovery, the log-normal model has become a key component of most time-dependent PSHA procedures and is routinely used by the Working Group on California Earthquake Probabilities (WGCEP, 1995).

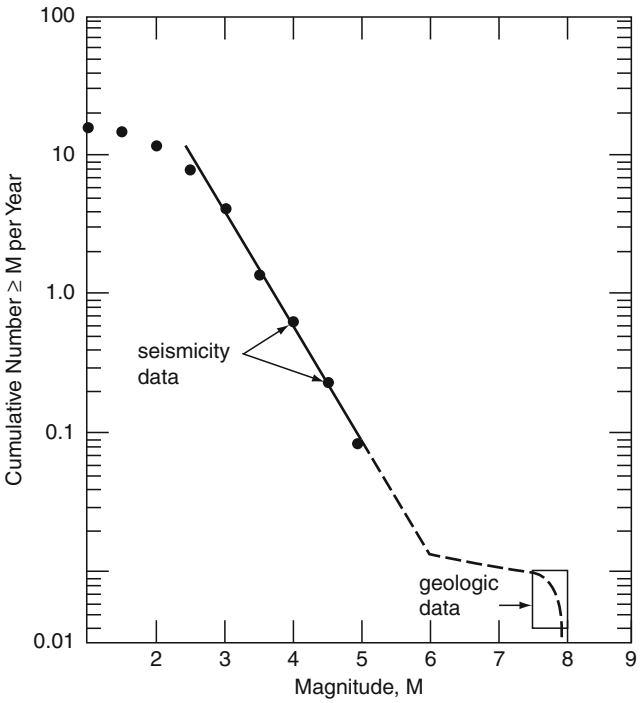
A time-dependent earthquake occurrence model which is applied more often is the Brownian passage time (BPT) distribution, also known as the inverse Gaussian distribution (Matthews et al., 2002). The model is described by two parameters: μ and σ, which, respectively, represent the mean time interval between the consecutive earthquakes and the standard deviation. The aperiodicity of earthquake occurrence is controlled by the variation coefficient $\alpha = \sigma/\mu$. For a small α, the aperiodicity of earthquake occurrence is small and the shape of distribution is almost symmetrical. For a large α, the shape of distribution is similar to log-normal model, i.e., skewed to the right and peaked at a smaller value than the mean. The straightforward control of aperiodicity of earthquake occurrence, by parameter α, makes the BPT model very attractive. It has been used to model earthquake occurrence in many parts of the world and has been applied by the Working Group on California Earthquake Probabilities (WGCEP, 1995).

Comparison of time-dependent with time-independent earthquake occurrence models have shown that the time-independent (Poissonian) model can be used for most engineering computations of PSHA. The exception to this rule is when the seismic hazard is dominated by a single seismic source, with a significant component of characteristic occurrence when the time interval from the last earthquake exceeds the mean time interval between consecutive events. Note that, in most cases, the information on strong seismic events provided by current databases is insufficient to distinguish between different models. The use of non-Poissonian models will therefore only be justified if more data will be available.

Alternative frequency-magnitude models. In the classic Cornell–McGuire procedure for PSHA assessment, it is assumed that earthquake magnitudes follows the Gutenberg–Richter relation truncated from the top by a seismic source characteristic, the maximum possible earthquake magnitude m_{max}. The PDF of this distribution is given by Equation 5.

Despite the fact that in many cases the Gutenberg–Richter relation describes magnitude distributions within seismic source zones sufficiently well, there are some instances where it does not apply. In many places, especially for areas of seismic belts and large faults, the Gutenberg–Richter relation underestimates the occurrence of large magnitudes. The continuity of the distribution (Equation 5) breaks down. The distribution is adequate only for small events up to magnitude 6.0–7.0. Larger events tend to occur within a relatively narrow range of magnitudes (7.5–8.0), but with a frequency higher than that predicted by the Gutenberg–Richter relation. These events are known as *characteristic earthquakes* (Youngs and Coppersmith, 1985, Figure 4). Often it is assumed that characteristic events follow a truncated Gaussian magnitude distribution (WGCEP, 1995).

There are several alternative frequency-magnitude relations that are used in PSHA. The best known is probably the relation by Merz and Cornell (1973), which accounts for a possible curvature in the log-frequency-magnitude relation (1) by the inclusion of a quadratic term of magnitude. Departure from linearity of the distribution (Equation 1) is built into the model by Lomnitz-Adler and Lomnitz (1979). The model is based on simple physical considerations of strain accumulation and release at plate boundaries. Despite the fact that m_{max} is not present in the model, it provides estimates of the occurrence of large events which are more realistic than those predicted by the Gutenberg–Richter relation (1). When seismic hazard is caused by induced seismicity, an alternative distribution to the Gutenberg–Richter model (1) is always required. For example, the magnitude distributions of tremors generated by mining activity are multimodal and change their shape in time (Gibowicz and Kijko, 1994).



Seismic Hazard, Figure 4 Gutenberg–Richter characteristic earthquake magnitude distribution. The model combines frequency-magnitude Gutenberg–Richter relation with a uniform distribution of characteristic earthquakes. The model predicts higher rates of exceedance at magnitudes near the characteristic earthquake magnitude. (After Youngs and Coppersmith, 1985).

Often, the only possible method that can lead to a successful PSHA for mining areas is the replacement of the analytical, parametric frequency-magnitude distribution by its model-free, non-parametric counterpart (Kijko et al., 2001).

Two more modifications of the recurrence models are regularly introduced: one when earthquake magnitudes are uncertain and the other when the seismic occurrence process is composed of temporal trends, cycles, short-term oscillations, and pure random fluctuations. The effect of error in earthquake magnitude determination (especially significant for historic events) can be minimized by the simple procedure of correction of the earthquake magnitudes in a catalogue (e.g., Rhoades, 1996). The modelling of random fluctuations in earthquake occurrence is often done by introducing compound distributions in which parameters of earthquake recurrence models are treated as random variables (Campbell, 1982).

Ground motion prediction equations

The assessment of seismic hazard at a site requires knowledge of the prediction equation of the particular strong motion parameter, as a function of distance, earthquake magnitude, faulting mechanism, and often the local site condition below the site. The most simple and most commonly used form of a prediction equation is

$$\ln(y) = c_1 - c_2 m - c_3 \ln(r) - c_4 r + c_5 F + c_6 S + \varepsilon, \tag{19}$$

where y is the amplitude of the ground motion parameter (PGA, MM intensity, seismic record duration, spectral acceleration, etc.); m is the earthquake magnitude; r is the shortest earthquake distance from the site to the earthquake source; F is responsible for the faulting mechanism; S is a term describing the site effect; and ε is the random error with zero mean and standard deviation $\sigma_{\ln(y)}$, which has two components: epistemic and aleatory.

The coefficients $c_1, \ldots, c_6$ are estimated by the least squares or maximum likelihood procedure, using strong motion data. It has been found that the coefficients depend on the tectonic settings of the site. They are different for sites within stable continental regions, active tectonic regions, or subduction zone environments. Assuming that $\ln(y)$ has a normal distribution, regression of (Equation 19) provides the mean value of $\ln(y)$, the exponent of which corresponds to the median value of y, $\breve{y}$. Since the log-normal distribution is positively skewed, the mean value of y, $\bar{y}$, exceeds the median value $\breve{y}$ by a factor of $\exp(-0.5\sigma^2_{\ln(y)})$. This indicates that the seismic hazard for a particular site is higher when expressed in terms of $\bar{y}$, than the hazard for the same site expressed in terms of $\breve{y}$. It has been shown that the ground motion prediction equation remains a particularly important component of PSHA since its uncertainty is a major contributor to uncertainty of the PSHA results (SSHAC, 1997).

Uncertainties in PSHA

Contemporary PSHA distinguishes between two types of uncertainties: aleatory and epistemic.

The *aleatory uncertainty* is due to randomness in nature; it is the probabilistic uncertainty inherent in any random phenomenon. It represents unique details of any earthquake as its source, path, and site and cannot be quantified before the earthquake occurrence and cannot be reduced by current theories, acquiring addition data or information. It is sometimes referred as "randomness," "stochastic uncertainty," or "inherent variability" (SSHAC, 1997) and is denoted as U_R (McGuire, 2004). The typical examples of aleatory uncertainties are: the number of future earthquakes in a specified area; parameters of future earthquakes such as origin times, epicenter coordinates, depths and their magnitudes; size of the fault rupture; associated stress drop and ground motion parameters like PGA, displacement or seismic record duration at the given site. The aleatory uncertainties are characteristic to the current model and cannot be reduced by the incorporation of addition data. It can only be reduced by the conceptualization of a better model.

The *epistemic uncertainty*, denoted as U_K is the uncertainty due to insufficient knowledge about the model or its parameters. The model (in the broad sense of its meaning; as, e.g., a particular statistical distribution) may be approximate and inexact, and therefore predicts values that differ from the observed values by a fixed, but unknown, amount. If uncertainties are associated with numerical values of the parameters, they are also epistemic by nature. Epistemic uncertainty can be reduced by incorporating additional information or data. Epistemic distributions of a model's parameters can be updated using the Bayes' theorem. When new information about parameters is significant and accurate, these epistemic distributions of parameters become delta functions about the exact numerical values of the parameters. In such a case, no epistemic uncertainty about the numerical values of the parameters exists and the only remaining uncertainty in the problem is aleatory uncertainty.

In the past, epistemic uncertainty has been known as statistical or professional uncertainty. The examples of the epistemic uncertainties are: boundaries of seismic sources, distributions of seismic sources parameters (e.g., annual rate of seismic activity λ, b-value and m_{max}), or median value of the ground motion parameter given the source properties.

Aleatory uncertainties are included in the PSHA by means of integration (Equation 5) and they are represented by the hazard curve. In contrast, epistemic uncertainties are included through the use of an alternative hypothesis – different sets of parameters with different numerical values, different models, or through a *logic tree*. Therefore, by default, if in the process of PSHA, the logic tree formalism is applied, the resulting uncertainties of the hazard curve are of epistemic nature.

The major benefit of the separation of uncertainties into aleatory and epistemic is potential guidance in the preparation of input for PSHA and the interpretation of the results. Unfortunately, the division of uncertainties into aleatory and epistemic is model dependent and to a large extent arbitrary, indefinite, and confusing (Panel of Seismic Hazard Evaluation, 1997).

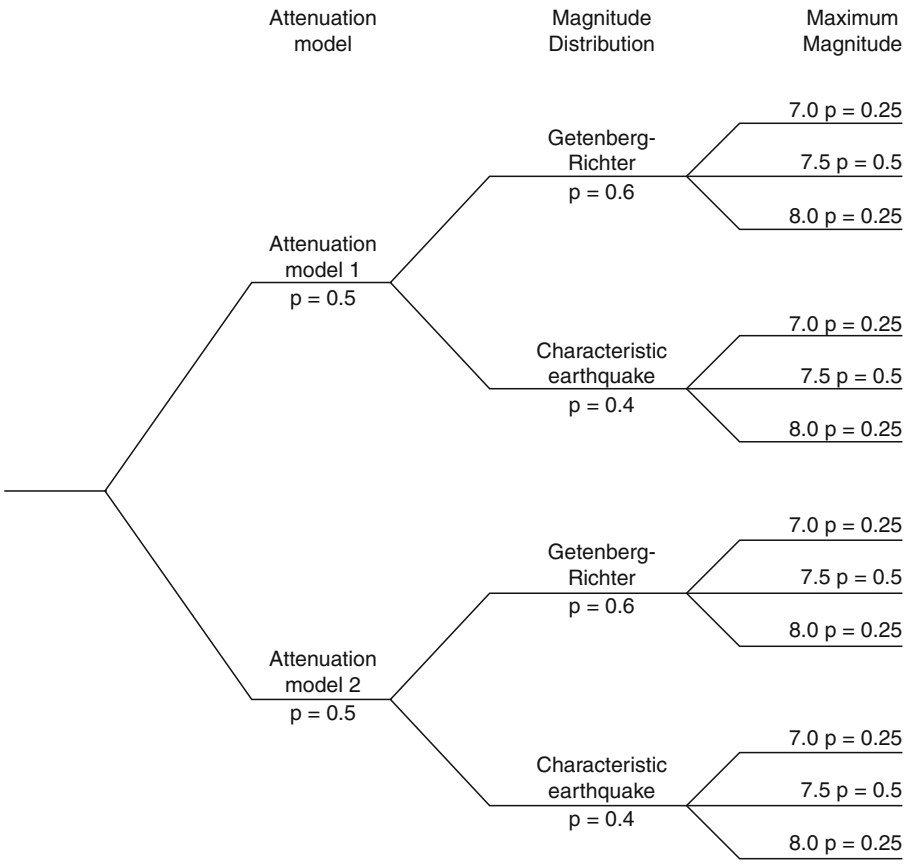
Logic tree

The mathematical formalism of PSHA computation, (Equations 7 and 9), integrates over all random (aleatory) uncertainties of a particular seismic hazard model. In many cases, however, because of our lack of understanding of the mechanism that controls earthquake generation and wave propagation processes, the best choices for elements of the seismic hazard model is not clear. The uncertainty may originate from the choice of alternative seismic sources, competitive earthquake recurrence models and their parameters, as well as from the choice of the most appropriate ground motion. The standard approach for the explicit treatment of alternative hypotheses, models, and parameters is the use of a *logic tree*. The logic tree formalism provides a convenient tool for quantitative treatment of any alternatives. Each node of the logic tree (Figure 5) represents uncertain assumptions, models, or parameters, and the branches extending from each node are the discrete uncertainty alternatives.

In the logic tree analysis, each branch is weighted according to its probability of being correct. As a result, each end branch represents a hazard curve with an assigned weight, where the sum of weights of all the hazard curves is equal to 1. The derived hazard curves are thus used to compute the final (e.g., mean) hazard curve and their confidence intervals. An example of a logic tree is shown in Figure 5. The alternative hypotheses account for uncertainty in the ground motion attenuation model, the magnitude distribution model and the assigned maximum magnitude m_{max}.

Controversy

Despite the fact that the PSHA procedure, as we know it in its current form, was formulated almost half of century ago, it is not without controversy. The controversy surrounds questions such as: (1) the absence of the upper limit of ground motion parameters, (2) division of uncertainties between aleatory and epistemic, and (3) methodology itself, especially the application of the logic tree formalism.



Seismic Hazard, Figure 5 An example of a simple logic tree. The alternative hypothesis accounts for uncertainty in ground motion attenuation relation, magnitude distribution model, and the assigned maximum magnitude m_{max}.

In most currently used Cornell–McGuire-based PSHA procedures, the ground motion parameter used to describe the seismic hazard is distributed log-normally. Since the log-normal distribution is unlimited from the top, it results in a nonzero probability of unrealistically high values for the ground motion parameter, e.g., PGA ≈ 20g, obtained originally from a PSHA for a nuclear-waste repository at Yucca Mountain in the USA (Corradini, 2003). The lack of the upper bound of earthquake-generated ground motion in current hazard assessment procedures has been identified as the "missing piece" of the PSHA procedure (Bommer et al., 2004).

Another criticism of the current PSHA procedure concerns portioning of uncertainties into aleatory and epistemic. As noted in the section (Uncertainties in PSHA) above, the division between aleatory and epistemic uncertainty remains an open issue.

A different criticism comes from the ergodic assumptions which underlie the formalism of the PSHA procedure. The ergodic process is a random process in which the distribution of a random variable in space is the same as distribution of that variable at a single point when sampled as a function of time (Anderson and Brune, 1999). It has been shown that the major contribution to PSHA uncertainty comes from uncertainty of the ground motion prediction equation. The uncertainty of the ground motion parameter y, is characterized by its standard deviation, $\sigma_{\ln(y)}$, which is calculated as the misfit between the observed and predicted ground motions at several seismic stations for a small number of recorded earthquakes.

Thus, $\sigma_{\ln(y)}$ mainly characterizes the spatial and not the temporal uncertainty of ground motion at a single point. This violates the ergodic assumption of the PSHA procedure. According to Anderson and Brune (1999), such violation leads to overestimation of seismic hazard, especially when exposure times are longer than earthquake return times. In addition, Anderson et al. (2000) shows that high-frequency PGAs observed at short distances do not increase as fast as predicted by most ground motion relations. Therefore, the use of the current ground motion prediction equations, especially relating to seismicity recorded at short distances, results in overestimation of the seismic hazard.

A similar view has been expressed by Wang and Zhou (2007) and Wang (2009). Inter alia, they argue that in the Cornell–McGuire-based PSHA procedure, the ground motion variability is not treated correctly. By definition, the ground motion variability is implicitly or explicitly dependent on earthquake magnitude and distance; however, the current PSHA procedure treats it as an independent random variable. The incorrect treatment of ground motion variability results in variability in earthquake magnitudes and distance being counted twice. They conclude that the current PSHA is not consistent with modern earthquake science, is mathematically invalid, can lead to unrealistic hazard estimates, and causes confusion. Similar reservations have been expressed in a series of papers by Klügel (Klügel, 2007 and references therein).

Equally strong criticism of the currently PSHA procedure has been expressed by Castanos and Lomnitz (2002). The main target of their criticism is the logic tree, the key component of the PSHA. They describe the application of the logic tree formalism as a misunderstanding in probability and statistics, since it is fundamentally wrong to admit "expert opinion as evidence on the same level as hard earthquake data."

The science of seismic hazard assessment is thus subject to much debate, especially in the realms where instrumental records of strong earthquakes are missing. At this time, PSHA represents a best-effort approach by our species to quantify an issue where not enough is known to provide definitive results, and by many estimations a great deal more time and measurement will be needed before these issues can be resolved.

Further reading: There are several excellent studies that describe all aspects of the modern PSHA. McGuire (2008) traces the intriguing historical development of PSHA. Hanks and Cornell (1999), and Field (1995) present an entertaining and unconventional summary of the issues related to PSHA, including its misinterpretation. Reiter (1990) comprehensively describes both the deterministic as well as probabilistic seismic hazard procedures from several points of view, including a regulatory perspective. Seismic hazard from the geologist's perspective is described in the book by Yeats et al. (1997). Kramer (1996) provides an elegant, coherent, and understandable description of the mathematical aspects of both DSHA and PSHA. Anderson et al. (2000), Gupta (2002), and Thenhaus and Campbell (2003) present excellent overviews covering theoretical, methodological, as well as procedural issues of modern PSHA. Finally, the most comprehensive treatment to date of all aspects of PSHA, including treatment of *aleatory* and *epistemic* uncertainties, is provided by the SSHAC – Senior Seismic Hazard Committee (1997) – report and in book form by McGuire (2004). The presentations here benefited from all quoted above sources, especially the excellent book by Kramer (1996).

Summary

Seismic hazard is a term referring to any physical phenomena associated with an earthquake (e.g., ground motion, ground failure, liquefaction, and tsunami) and their effects on land, man-made structures, and socioeconomic systems that have the potential to produce a loss. The term is also used, without regard to a loss, to indicate the probable level of ground shaking occurring at a given point within a certain period of time. Seismic hazard analysis is an expression referring to quantification of the expected ground motion at the particular site. Seismic hazard analysis can be performed deterministically, when a particular earthquake scenario is considered, or probabilistically, when the likelihood or frequency of a specified level of ground motion at a site during a specified exposure time is evaluated. In principle, any natural hazard caused by

seismic activity can be described and quantified in terms of the probabilistic methodology. Classic probabilistic seismic hazard analysis (PSHA) includes four steps: (1) identification and parameterization of the seismic sources, (2) specification of temporal and magnitude distributions of earthquake occurrence, (3) calculation of ground motion prediction equations and their uncertainty, and (4) integration of uncertainties in earthquake location, earthquake magnitude, and ground motion prediction equations into the hazard curve.

An integral part of PSHA is the assessment of uncertainties. Contemporary PSHA distinguishes between two types of uncertainties: aleatory and epistemic. The aleatory uncertainty is due to randomness in nature; it is the probabilistic uncertainty inherent in any random phenomenon. The aleatory uncertainties are characteristic to the current model and cannot be reduced by the incorporation of addition data. The epistemic uncertainty is the uncertainty due to insufficient knowledge about the model or its parameters. Epistemic uncertainty can be reduced by incorporating additional information or data. Aleatory uncertainties are included in the probabilistic seismic hazard analysis due to the integration over these uncertainties, and they are represented by the hazard curve. In contrast, epistemic uncertainties are included through the use of alternative models, different sets of parameters with different numerical values or through a logic tree.

Unfortunately, the PSHA procedure, as we know it in its current form, is not without controversy. The controversy arises from questions such as: (1) the absence of the upper limit of ground motion parameter, (2) division of uncertainties between aleatory and epistemic, and (3) methodology itself, especially the application of the logic tree formalism.

Bibliography

Aki, K., 1965. *Maximum Likelihood Estimate of b in the Formula logN = a−bM and its Confidence Limits*. University of Tokyo: Bulletin Earthquake Research Institute, Vol. 43, pp. 237–239.

Anderson, J. G., and Brune, J. N., 1999. Probabilistic seismic hazard analysis without the ergodic assumptions. *Seismological Research Letters*, **70**, 19–28.

Anderson, J. G., Brune, J. N., Anooshehpoor, R., and Shean-Der, Ni., 2000. New ground motion data and concepts in seismic hazard analysis. *Current Science, Special Section: Seismology*, **79**, 1278–1290.

Bender, B., 1984. Incorporation acceleration variability into seismic hazard analysis. *Bulletin Seismological Society of America*, **74**, 1451–1462.

Bommer, J. J., Abrahamson, N. A., Strasser, F. O., Pecker, A., Bard, P. Y., Bugnum, H., Cotton, F., Fäh, D., Sabette, F., Scherbaum, F., and Studer, J., 2004. The challenge of defining upper bounds on earthquake ground motions. *Seismological Research Letters*, **75**, 82–95.

Campbell, K. W., 1982. Bayesian analysis of extreme earthquake occurrences. Part I. Probabilistic hazard model. *Bulletin Seismological Society of America*, **72**, 1689–1705.

Castanos, H., and Lomnitz, C., 2002. PSHA: is it science? opinion paper. *Engineering Geology*, **66**, 315–317.

Cooke, P., 1979. Statistical inference for bounds of random variables. *Biometrika*, **66**, 367–374.

Cornell, C. A., 1968. Engineering seismic risk analysis. *Bulletin Seismological Society of America*, **58**, 1583–1606.

Cornell, C. A., 1971. Bayesian Statistical Decision Theory and Reliability Based Design. In Freudenthal, A. M. (ed.), *Proceedings of the International Conference on Structural Safety and Reliability*, April 9-11, 1969. Smithsonian Institute: Washington DC, pp. 47–66.

Cornell, C. A., 1994. Statistical analysis of maximum magnitudes. In Johnston, A. C., Coppersmith, K. J., Kanter, L. R., and Cornell, C. A. (eds.), *The Earthquakes of Stable Continental Regions*. Palo Alto: Electric Power Research Institute. Assessment of Large Earthquake Potential, Vol. 1, pp. 5–1, 5–27.

Corradini, M. L., 2003. Letter from chairman of the US Nuclear Waste Technical Review Board to the director of the Office of Civilian Radioactive Waste Management; available at: http://www.nwtrb.gov/corr/mlc010.pdf

Cramer, C. H., Petersen, M. D., Cao, T., Toppozada, T. R., and Reichle, M., 2000. A time dependent probabilistic seismic hazard model for California. *Bulletin Seismological Society of America*, **90**, 1–21.

Epstein, B., and Lomnitz, C., 1966. A model for the occurrence of large earthquakes. *Nature*, **211**, 954–956.

Field, D. H., 1995. Probabilistic seismic hazard analysis. A primer. (http://www.relm.org.tutorial_materials).

Frohlich, C., 1998. Does maximum earthquake size depend on focal depth? *Bulletin Seismological Society of America*, **88**, 329–336.

Giardini, D., 1999. The Global Seismic hazard Assessment Program (GSHAP) 1992–1999. *Annali di Geofisica*, **42**, 957–1230.

Gibowicz, S. J., and Kijko, A., 1994. *An Introduction to Mining Seismology*. San Diego: Academic.

Gupta, L. D., 2002. The state of the art in seismic hazard analysis. *ISET Journal of Earthquake Technology*, **39**, 311–346. Paper No. 428.

Gutenberg, B., and Richter, C. F., 1944. Frequency of earthquakes in California. *Bulletin Seismological Society of America*, **34**, 185–188.

Hanks, T. C., and Cornell, C. A., 1999. Probabilistic Seismic Hazard Analysis: A Beginners Guide (available from T.C. Hanks at thanks@usgs.gov).

Jin, A., and Aki, K., 1988. Spatial and temporal correlation between coda Q and seismicity in China. *Bulletin Seismological Society of America*, **78**, 741–769.

Kijko, A., 2004. Estimation of the maximum earthquake magnitude m_{max}. *Pure and Applied Geophysics*, **161**, 1–27.

Kijko, A., 2008. *Data Driven Probabilistic Seismic Hazard Assessment Procedure for Regions with Uncertain Seimogenic Zones*. NATO Monograph, Husebye E. S. (ed.), Earthquake monitoring and seismic hazard mitigation. Seismic hazard analysis and assessment. Springer Science, Chap. 16, pp. 237–251.

Kijko, A., and Graham, G., 1999. Parametric-historic. Procedure for probabilistic seismic hazard analysis. Part II: assessment of seismic hazard at specified site. *Pure and Applied Geophysics*, **154**, 1–22.

Kijko, A., and Sellevoll, M. A., 1992. Estimation of earthquake hazard parameters from incomplete data files, part II, incorporation of magnitude heterogeneity. *Bulletin Seismological Society of America*, **82**, 120–134.

Kijko, A., Lasocki, S., and Graham, G., 2001. Nonparametric seismic hazard analysis in mines. *Pure and Applied Geophysics*, **158**, 1655–1675.

Klügel, J.-U., 2007. Error inflation in probabilistic seismic hazard analysis. *Engineering Geology*, **90**, 186–192.

Kramer, S. L., 1996. *Geotechnical Earthquake Engineering*. Englewood Cliffs: Prentice-Hill.

Lomnitz-Adler, J., and Lomnitz, C., 1979. A modified form of the Gutenberg-Richter magnitude-frequency relation. *Bulletin Seismological Society of America*, **69**, 1209–1214.

Main, I. G., and Burton, P. W., 1984. Information theory and the earthquake frequency-magnitude distribution. *Bulletin Seismological Society of America*, **74**, 1409–1426.
Matthews, M. V., Ellsworth, W. L., and Reasenberg, P. A., 2002. A Brownian model for recurrent earthquakes. *Bulletin Seismological Society of America*, **92**, 2233–2250.
McCalpin, J. P. (ed.), 1996. *Paleoseimology*. New York: Academic.
McGarr, A., 1984. Some applications of seismic source mechanism studies to assessing underground hazard. In Gay, N. C., and Wainwright, E. H. (eds.), *Rockburst and Seismicity in Mines*. Johannesburg: South African Institute of Mining and Metallurgy (Symposium Series No. 6, pp. 199–208).
McGuire, R. K., 2004. *Seismic Hazard and Risk Analysis*. Oakland: Earthquake Engineering Research Institute, MNO-10.
McGuire, R. K., 2008. Review. Probabilistic seismic hazard analysis: early history. *Earthquake Engineering and Structural Dynamics*, **37**, 329–338.
Merz, H. A., and Cornell, C. A., 1973. Seismic risk based on quadratic magnitude frequency low. *Bulletin Seismological Society of America*, **69**, 1209–1214.
Muir-Wood, R., 1993. From global seismotectonics to global seismic hazard. *Annali di Geofisica*, **36**, 153–168.
Nishenko, S. P., and Buland, R., 1987. A generic recurrence interval distribution for earthquake forecasting. *Bulletin Seismological Society of America*, **77**, 1382–1399.
Nuttli, O. W., 1981. *On the Problem of Maximum Magnitude of Earthquakes*. U.S. Geological Survey Open-file Report 81:1–13.
Ogata, Y., 1999. Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes. *Journal of Geophysical Research*, **104**(17), 995–18014.
Panel of Seismic Hazard Analysis, 1997. *Review of Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*. Washington DC: National Academy.
Petersen, M. D., Frankel, A. D., Harmsen, S. C., Mueller, C. S., Haller, K. M., Wheeler, R. L., Wesson, R. L., Zeng, Y., Boyd, O. S., Perkins, D. M., Luco, N., Field, E. H., Wills, C. J., and Rukstales, K. S., 2008. Documentation for the 2008 update of the United States national seismic hazard maps. *U.S. Geological Survey Open-File Report 2008–1128*, 61 p.
Pisarenko, V. F., Lyubushin, A. A., Lysenko, V. B., and Golubieva, T. V., 1996. Statistical estimation of seismic hazard parameters: maximum possible magnitude and related parameters. *Bulletin Seismological Society of America*, **86**, 691–700.
Reiter, L., 1990. *Earthquake Hazard Analysis: Issues and Insights*. New York: Columbia University Press.
Rhoades, D. A., 1996. Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties. *Tectonophysics*, **258**, 71–83.
Rhoades, D., Van Dissen, R. J., and Dowrick, D. J., 1994. On the handling of uncertainties in estimating the hazard rupture on a fault. *Journal of Geophysical Research*, **99**, 13,701–13,712.
Shimazaki, K., and Nakata, T., 1980. Time-Predictable recurrence model for large earthquakes. *Geophysical Research Letters*, **7**, 279–282.
SSHAC - Senior Seismic Hazard Committee, 1997. *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*. NUREG/CR-6372, UCR-ID-122160, Main Report 1. Prepared for Lawrence Livermore National Laboratory.
Thenhaus, P. C., and Campbell, K. W., 2003. Seismic hazard analysis. In Chen, W. F., and Scawthorn, C. (eds.), *Earthquake Engineering Handbook*. Boca Raton: CRC Press, pp. 8-1–8-50.
Utsu, T., 1965. A method for determining the value of *b* on the formula log $n = a - bM$ showing the magnitude–frequency relation for earthquakes. *Geophysical Bulletin Hokkaido University*, **13**, 99–103. In Japan: Engl. abstr.
Veneziano, D., Cornell, C. A., and O'Hara, T., 1984. *Historic Method for Seismic Hazard Analysis*. Electric Power Research Institute, Report, NP-3438, Palo Alto.
Wang, Z., 2009. Comment on "Sigma: issues, insights, and challenges" by F. O. Strasser, N. A. Abrahamson, and J. J. Bommer. *Seismological Research Letters*, **80**, 491–493.
Wang, Z., and Zhou, M., 2007. Comment on "why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates?" by Julian J. Bommer, and Norman A. Abrahamson. *Bulletin Seismological Society of America*, **97**, 2212–2214.
Ward, S. N., 1997. More on M_{max}. *Bulletin Seismological Society of America*, **87**, 1199–1208.
Weichert, D. H., 1980. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin Seismological Society of America*, **70**, 1337–1346.
Wells, D. L., and Coppersmith, K. J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin Seismological Society of America*, **84**, 974–1002.
WGCEP (Working Group on Central California Earthquake Probabilities), 1995. Seismic hazard in southern California: probable earthquakes, 1994 to 2024. *Bulletin Seismological Society of America*, **85**, 379–439.
Wheeler, R. L., 2009. *Methods of Mmax estimation East of Rocky mountains*. USGS, Open-File Report 2009–1018.
Woo, G., 1996. Kernel estimation methods for seismic hazard area source modeling. *Bulletin Seismological Society of America*, **86**, 353–362.
Yeats, R. S., Sieh, K., and Allen, C. R., 1997. *The Geology of Earthquakes*. New York: Oxford University Press.
Youngs, R. R., and Coppersmith, K. J., 1985. Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bulletin Seismological Society of America*, **75**, 939–964.

Cross-references

Characteristic Earthquakes and Seismic Gaps
Earthquake, Magnitude
Earthquakes, Early and Strong Motion Warning
Earthquakes, Intensity
Earthquakes, Shake Map
Earthquakes, Strong-Ground Motion
Seismic Zonation
Statistical Seismology

SEISMIC IMAGING, OVERVIEW

Gerard T. Schuster
Division of Environmental and Earth Sciences, King Abdullah University of Science and Technology, Thule, Saudi Arabia

Definition

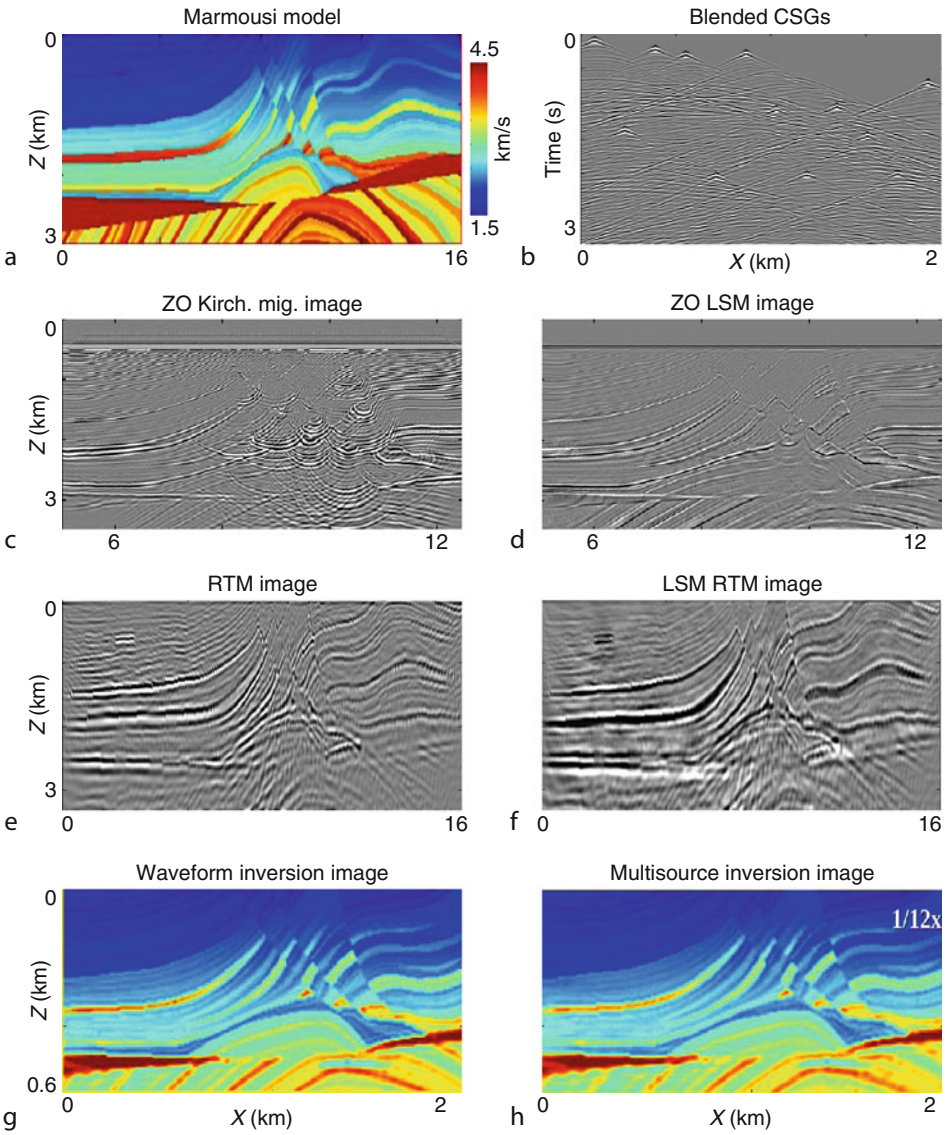
Seismic imaging or tomography (tomo = slice and graph = picture) is a procedure for estimating the earth's rock parameters from seismic data. These rock parameters can be represented by the spatial distribution of, e.g., P-wave velocity, S-wave velocity, porosity, density, or anisotropic parameters. The result of inversion is graphically presented as a 2-D or 3-D grid of pixels, where each pixel

contains the value of the model parameter of interest, which is P velocity in Figure 1g–h. Such tomograms are used to estimate the geometry and lithology of geologic layers, and can help exploration geophysicists and earthquake seismologists understand the evolution of the earth's interior.

There are five main types of seismic imaging: seismic migration, least squares migration, full waveform inversion (FWI), phase-like inversion, and migration velocity analysis. Four of these methods can be derived as special cases of finding the optimal model that minimizes a waveform or a phase-related misfit function. This entry will present these methods as they are applied to body wave events in seismic data, and overview the current best practices in waveform tomography.

Inversion theory

There are four steps to inverting for the model **m** from seismic data **d**. We will assume that the input data are either seismic traces, generated by man-made or earthquake sources, or some skeletonized part of the traces, such as first arrival traveltimes or the phase of an event at some frequency.



Seismic Imaging, Overview, Figure 1 (**a**) Marmousi velocity model, (**b**) time-shifted and blended shot gathers, (**c**) zero-offset Kirchhoff migration image, (**d**) zero-offset least squares migration image after 30 iterations (courtesy of Naoshi Aoki), (**e**) reverse time migration (RTM) image, (**f**) least squares RTM image (Dai and Schuster, 2009), (**g**) waveform tomogram after 50 iterations using the prestack gathers, and (**h**) waveform tomogram where the inputs are eight supergathers, each with 12 phase-encoded shot gathers (Zhan and Schuster, in press); images are at different scales from one another.

Step 1: $\mathbf{L}(\mathbf{m}) = \mathbf{d}$. Establish the mathematical relationship $\mathbf{L}(\mathbf{m}) = \mathbf{d}$ between the seismic data $\mathbf{d}$ and model $\mathbf{m}$.

$$\mathbf{d} = \mathbf{L}(\mathbf{m}), \tag{1}$$

where $\mathbf{L}$ represents the forward modeling operator for the actual model. Equations A1–A3 in Table 1 show the three steps in establishing a linearized version of Equation 1 for the acoustic wave equation, where the extensions to the elastic (Mora, 1987) and anisotropic (Barnes et al., 2008; Operto et al., 2009) wave equations are tedious but straightforward.

Step 2: Discretize $\mathbf{m}$, $\mathbf{d}$, and $\mathbf{L}$. Discretize the 3D earth model into a 3D grid of N physical parameters (e.g., unknown slowness in each cell) and assemble the unknowns into the $N \times 1$ vector $\mathbf{m}$. Discretize the seismic traces in both space and time (or frequency) into an $M \times 1$ vector $\mathbf{d}$ of data measurements. In this case, $\mathbf{L}$ reduces to a $M \times N$ matrix. The forward modeling for FWI requires knowledge of the source wavelet, which can be estimated in a variety of ways: stacking of the direct arrival at different near-offset hydrophones in marine data, iterative inversion of the source wavelet (Mora, 1987; Zhou et al., 1997; Pratt, 1999), or deconvolution of the shot gather with a time-windowed near-offset trace (Sheng, personal communication). The time window can be a few periods long centered about the direct arrival.

Step 3. Linearize $\mathbf{L}\delta\mathbf{m} \approx \delta\mathbf{d}$. Linearize the nonlinear relationship between the data and model. Expanding the ith data measurement $d_i(\mathbf{m})$ to first order in $\delta\mathbf{m}$ by a Taylor series about a first-guess model $\mathbf{m}_o$ (close to the true model) gives the linearized estimate:

$$\begin{aligned} d_i(\mathbf{m}) &\approx d_i(\mathbf{m}_o) + \sum_j \frac{\delta d_i(\mathbf{m}_o)}{\delta m_j}\,\delta m_j \rightarrow \delta d_i(\mathbf{m}) \\ &= \sum_j \overbrace{\left[\partial d_i(\mathbf{m}_o)/\partial m_j\right]}^{\text{wavepath function}}\,\delta m_j, \end{aligned} \tag{2}$$

or in matrix-vector notation

$$\delta\mathbf{d} = \mathbf{L}\delta\mathbf{m}. \tag{3}$$

Here, $\partial d_i(\mathbf{m}_o)/\partial m_j = \lim_{\Delta m_j \rightarrow 0}[d_i(\mathbf{m}_o + \hat{\mathbf{j}}\Delta m_j) - d_i(\mathbf{m}_o)]/\Delta m_j$ is the Fréchet derivative with respect to the jth parameter; the data residual $\delta d_i = [d_i(\mathbf{m}) - d_i(\mathbf{m}_o)]$ is the difference between the ith components of the predicted data vector $\mathbf{d}(\mathbf{m}_o)$ and the observed data vector $\mathbf{d}(\mathbf{m})$; and $\hat{\mathbf{j}}$ is the jth unit vector in the finite-dimensional model space. The model perturbation $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_o$ is the difference between the actual model $\mathbf{m}$ and the guessed model $\mathbf{m}_o$, and $\mathbf{L}$ is now interpreted as the Jacobian matrix. Its elements $[\mathbf{L}]_{ij} = \partial d_i(\mathbf{m}_o)/\partial m_j$ determine the sensitivity of the data to the perturbations in the model. For a windowed arrival, the Jacobian plots out in model space as a wavepath for a single source-receiver pair (Woodward, 1992; Luo, 1992; Dahlen et al., 2002; Marquering et al., 2002; Montelli et al., 2004; van der Hilst and de Hoop, 2006; Xu and Xie, 2009); and the velocity variations within its first Fresnel zone mostly influence the event of interest. Equation A3 shows that the linearized equations take the form of the Born approximation (Stolt and Benson, 1986) to the

Seismic Imaging, Overview, Table 1 Symbols and specific formulas for inverting the Helmholtz equation for the perturbed slowness distribution $\delta s(\mathbf{x})$, where $\delta s(\mathbf{x})$ is assumed to be small compared to the background slowness model $s(\mathbf{x})$. The harmonic pressure field is denoted by $p(\mathbf{x}|\mathbf{s})$ for a source localized about $\mathbf{s}$, and the Green's function for the Helmholtz equation is given by $G(\mathbf{x}|\mathbf{s})$ for a point source at $\mathbf{s}$ and observer at $\mathbf{x}$. The source wavelet spectrum is $W(\omega)$ at the angular frequency ω and the body force term localized about the point $\mathbf{s}$ is denoted by $f(\mathbf{x}|\mathbf{s})$. The predicted (observed) traveltime of an event received at $\mathbf{g}$ for a source at $\mathbf{s}$ is denoted as $\tau(\mathbf{g}|\mathbf{s})$ ($\tau(\mathbf{g}|\mathbf{s})^{\text{obs.}}$); and $\varepsilon = ||\mathbf{d} - \mathbf{d}^{\text{obs.}}||^2$ is the misfit function for waveform inversion

Symbol	Mathematical formula
(A1) Helmholtz equation: $\mathbf{A}(\mathbf{d},\mathbf{m}) = \mathbf{f}$	$(\nabla^2 + \omega^2 s(\mathbf{x})^2)p(\mathbf{x}\|\mathbf{s}) = f(\mathbf{x}\|\mathbf{s})$ $\mathbf{m} \rightarrow s(\mathbf{x}); \mathbf{d} \rightarrow p(\mathbf{x}\|\mathbf{s})$
(A2) Linearized Helmholtz equation: $\delta\mathbf{A}(\mathbf{d},\mathbf{m}) = 0$	$\left(\nabla^2 + \omega^2 s(\mathbf{x})^2\right)\delta p(\mathbf{x}\|\mathbf{s}) = -2\omega^2 s(\mathbf{x})\delta s(\mathbf{x}) p(\mathbf{x}\|\mathbf{s})$ $\delta\mathbf{m} \rightarrow \delta s(\mathbf{x}); \delta\mathbf{d} \rightarrow \delta p(\mathbf{x}\|\mathbf{s})$
(A3) Lippmann–Schwinger equation with Born approx.: $\delta\mathbf{d} = \mathbf{L}\delta\mathbf{m}$	$\delta p(\mathbf{g}\|\mathbf{s}) = -2\omega^2 \int G(\mathbf{g}\|\mathbf{x}) s(\mathbf{x}) \delta s(\mathbf{x}) W(\omega) G(\mathbf{x}\|\mathbf{s}) \mathrm{d}x^3$ where $p(\mathbf{x}\|\mathbf{s}) = W(\omega)G(\mathbf{x}\|\mathbf{s})$
(A4) Jacobian or wavepath function: $\delta\mathbf{d}/\delta\mathbf{m}$	$\delta p(\mathbf{g}\|\mathbf{s})/\delta s(\mathbf{x}) = -2s(\mathbf{x})\omega^2 W(\omega) G(\mathbf{g}\|\mathbf{x}) G(\mathbf{x}\|\mathbf{s})$
(A5) Misfit gradient or reverse time migration or $\mathbf{L}^\dagger[\mathbf{L}\delta\mathbf{m} - \delta\mathbf{d}]$: $\delta\varepsilon/\delta\mathbf{m}$	$\delta\varepsilon/\delta s(\mathbf{x}) = \gamma \int G(\mathbf{g}\|\mathbf{x})^* \Delta d(\mathbf{g}\|\mathbf{s}) G(\mathbf{x}\|\mathbf{s})^* \mathrm{d}g\mathrm{d}s$ where $\gamma = -2\omega^2 s(\mathbf{x}) W(\omega)$ and $\Delta d(\mathbf{g}\|\mathbf{s}) = p(\mathbf{g}\|\mathbf{s}) - p(\mathbf{g}\|\mathbf{s})^{\text{obs.}}$
(A6) Least squares migration or linearized inversion: $\mathbf{m}^{(k)} - \delta\varepsilon/\delta\mathbf{m}$	$m(\mathbf{x})^{(k)} - \gamma \int G(\mathbf{g}\|\mathbf{x})^* \Delta d(\mathbf{g}\|\mathbf{s})^{(k)} G(\mathbf{x}\|\mathbf{s})^* \mathrm{d}g\mathrm{d}s$
(A7) Waveform inversion: $\mathbf{m}^{(k)} - \delta\varepsilon/\delta\mathbf{m}$	$m(\mathbf{x})^{(k)} - \gamma \int G^{(k)}(\mathbf{g}\|\mathbf{x})^* \Delta d(\mathbf{g}\|\mathbf{s})^{(k)} G^{(k)}(\mathbf{x}\|\mathbf{s})^* \mathrm{d}g\mathrm{d}s$
(A8) Wave equation traveltime inversion: $\mathbf{m}^{(k)} - \delta\varepsilon/\delta\mathbf{m}$	$m(\mathbf{x})^{(k)} - \gamma \int G^{(k)}(\mathbf{g}\|\mathbf{x})^* \Delta d(\mathbf{g}\|\mathbf{s})^{(k)} G^{(k)}(\mathbf{x}\|\mathbf{s})^* \mathrm{d}g\mathrm{d}s$ where $\Delta d(\mathbf{g}\|\mathbf{s}) \approx p(\mathbf{g}\|\mathbf{s})^{\text{obs.}} \left(\tau(\mathbf{g}\|\mathbf{s}) - \tau(\mathbf{g}\|\mathbf{s})^{\text{obs.}}\right)$

Lippmann–Schwinger equation, and its kernel is the Jacobian given by Equation A4.

For a single source and a single receiver in a smoothly increasing velocity medium, Equation A4 plots out as a curved "fat" ray (trace a ray that connects the source and receiver and honors Snell's law. Surround this ray with a "fat finite-frequency ray" such that the propagation of events from the source to receiver within the "fat" ray differs in traveltime by no more than 1/2 the period of the source wavelet. This fat ray region is that portion of the earth which mostly influences the traveltime of the event of interest) that connects the source and receiver points. This fat ray is denoted as a wavepath by Woodward (1992) and forms the basis of finite-frequency traveltime tomography. There seems to be a general (Montelli et al., 2004), but not a universal (van der Hilst and de Hoop, 2006), agreement that finite-frequency tomography can be superior to that of ray-based tomography. As an example, earthquake seismologists use these wavepaths (renamed as banana-doughnuts) with finite-frequency tomography to invert earthquake traveltimes and surface wave data for deep mantle velocity variations attributed to, for example, ascending plumes of rocks (Dahlen et al., 2002; Marquering et al., 2002; Montelli et al., 2006).

Step 4: Solve $\mathbf{L}\delta\mathbf{m} = \delta\mathbf{d}$ by an iterative gradient method. Equation 3 is typically an overdetermined, inconsistent, and poorly conditioned system of equations. Therefore, the solution we seek is the one that minimizes the sum of the data misfit $||\mathbf{Lm} - \mathbf{d}||_K^2$ and model penalty $\lambda^2||\mathbf{Cm}||_W^2$ functions in the K and W norms (Clinthorne et al., 1993):

$$\varepsilon = \frac{1}{2}||\mathbf{L}\delta\mathbf{m} - \delta\mathbf{d}||_K^2 + \frac{\lambda^2}{2}||\mathbf{C}\delta\mathbf{m}||_W^2. \tag{4}$$

where $||\mathbf{L}\delta\mathbf{m} - \delta\mathbf{d}||_K^2 = (\mathbf{L}\delta\mathbf{m} - \delta\mathbf{d})^\dagger\mathbf{K}(\mathbf{L}\delta\mathbf{m} - \delta\mathbf{d})$ and $\mathbf{K}$ is sometimes taken as the inverse of the data covariance matrix (Tarantola, 1987) that is real and symmetric. The penalty function (also known as the regularization term) is defined as $||\delta\mathbf{Cm}||_W^2 = (\mathbf{Cm})^\dagger\mathbf{W}(\mathbf{Cm})$, where the real symmetric matrix $\mathbf{W}$ might strongly weight certain regions in the model because they mostly influence the data. The matrix $\mathbf{C}$ might be a spatial second-derivative operator in one or several directions so as to encourage solutions with smoothly varying model parameters along selected directions.

A simplified steepest descent solution for the ith model parameter m_i is given by

$$\begin{aligned} m_i^{(k+1)} &= m_i^{(k)} - \alpha \overbrace{\frac{\delta\varepsilon}{\delta m_i}}^{\text{gradient Equation A5}} \rightarrow \mathbf{m}^{(k+1)} \\ &= \mathbf{m}^{(k)} - \overbrace{\alpha\mathbf{L}^\dagger\delta\mathbf{d}^{(k)}}^{\text{migration of data residual}}, \end{aligned} \tag{5}$$

where $\lambda = 0$; $\mathbf{K} = \mathbf{I}$; α is the step length; and preconditioning is used (Beydoun and Mendes, 1989; Clinthorne et al., 1993; Causse et al., 1995) to accelerate convergence. A sequence of models is generated until the data residual falls below some acceptable level. In practice, a preconditioned conjugate gradient method (Mora, 1987; Luo and Schuster, 1991; Epanomeritakis et al., 2008), a Gauss-Newton Krylov solver (Erlangga and Hermann, 2009), or a limited memory quasi-Newton method (Pratt et al., 1998; Plessix, 2006) for different implementations of the adjoint method (Plessix, 2009) is implemented; and sometimes a direct matrix solver is used to find the Hessian inverse if the problem size is small enough (Pratt and Goulty, 1991). Approximations to the Hessian by a systematic procedure can be found in Thierry et al. (1999). There is strong evidence (Crase et al., 1990; Brossier et al., 2010) that using the L_1 norm misfit function is noticeably more resistant to data noise than the L_2 norm misfit function, and the combination of the two is sometimes the best choice.

Five types of seismic imaging methods

Five types of seismic imaging methods and their resolution properties will be discussed (there are many seismic inversion methods, but the five types discussed here are often used in the geophysical community): migration (modest resolution of reflectivity), least squares migration (high resolution of reflectivity), full waveform inversion (high resolution of velocity), phase-like inversion (modest resolution of velocity), and migration velocity analysis (modest resolution of velocity). The high-resolution methods typically pay the price of increased computational cost and decreased robustness, compared to the moderate-resolution methods with relatively low cost and desirable robustness. For reflection migration and inversion, the spatial resolution limits are approximately defined by a generalized Radon transform analysis (Beylkin et al., 1985); and the resolution limits for ray-based transmission tomography (Williamson, 1991) can be estimated by considering the width of the transmission Fresnel zone.

Migration is the first iterate solution of Equation 5, and least squares migration, sometimes known as linearized inversion, is the final iterative solution where the operator $\mathbf{L}$ is not updated after each iteration; also, $\mathbf{m}(\mathbf{x})$ represents the reflectivity model at the position $\mathbf{x}$. In contrast to migration, the waveform inversion tomogram is the final iterative solution, where the velocity model $\mathbf{m}$ and $\mathbf{L}$ are updated after every iteration; waveform inversion falls under the class of nonlinear optimization methods. It can be shown under certain assumptions that waveform inversion reduces to either wave equation traveltime tomography (Luo and Schuster, 1991) or ray-based tomography. These last two methods are classified as phase-like inversion methods. Unlike minimizing the data misfit function in Equation 4, migration velocity analysis updates the velocity model to minimize a *model misfit function*, which is the normed difference between the predicted migration image and the actual migration image

in the, e.g., common image gather. Convergence problems associated with local minima in the misfit function are reduced by emphasizing *flatness* in the CIG misfit function (Symes and Carazone, 1991; Shen et al., 1993).

Migration

If the starting model is a smoothly varying velocity distribution that only generates the accurate first arrival, then the input data $\mathbf{d} - \mathbf{d}^{(0)}$ residual becomes the scattered data. In this case the desired model is the reflectivity distribution, which is similar to the slowness perturbation function $\delta s(\mathbf{x})$, and the background velocity model is the inverse of slowness $s(\mathbf{x})$. If a finite-difference solution to the wave equation, for example, is used to calculate $G(\mathbf{x}|\mathbf{s})$ and $G(\mathbf{g}|\mathbf{x})$ in Equation A5, then the first model update $\delta\mathbf{m}^{(1)}$ in Equation 5 is known as the reverse time migration image (Whitmore, 1983; McMechan, 1983). If a one-way wave equation method is used to generate the Green's function, then the migration method is a phase-shift or FX-type algorithm (Stolt and Benson, 1986; Claerbout, 1992; Etgen et al., 2009). A diffraction-stack migration method results if the Green's function is replaced by its asymptotic approximation $G(\mathbf{x}|\mathbf{s}) = A(\mathbf{x},\mathbf{s})e^{i\omega\tau_{xs}}$ (here, $A(\mathbf{x}, \mathbf{s})$ accounts for geometric spreading losses, and τ_{xs} is the first arrival time for a ray that connects the source at $\mathbf{s}$ with the observer at $\mathbf{x}$. A ray-tracing method can be used to compute these traveltimes for a sufficiently high frequency and smoothly varying medium [Bleistein et al., 2001]). An example of standard poststack migration is depicted in Figure 1c, where the velocity model is shown in Figure 1a. The migration image depicts the reflectivity distribution computed by a Kirchhoff migration method with an eikonal traveltime solver.

Seismic migration is also used to migrate teleseismic body waves processed to form so-called receiver functions. Examples include Ryberg and Weber (2000), Sheehan et al. (2000), and also Bostock et al. (2001) who used a ray-Born inversion approach. More recently, Nowack et al. (2007) applied the Gaussian beam migration approach of Hill (2001) to the migration of teleseismic body waves. Despite its widespread use and its robustness, migration is considered to be a moderate-resolution method because it approximates the inverse Hessian matrix $[\mathbf{L}^{\dagger}\mathbf{L}]^{-1}$ by a diagonal matrix.

Least squares migration

If the background model $1/s(\mathbf{x})$ is not updated after each iteration (i.e., $\mathbf{L}$ in Equation 5 is independent of the k index), then $\mathbf{m}^{(k+1)}$ for large k is known as the least squares migration image (Nemeth et al., 1999; Duquet et al., 2000). As in standard migration, the model to be iteratively updated is the reflectivity distribution and not the velocity model. The iterative least squares migration Equation A6 is interpreted as a sequence of standard migrations, where the data residual is backprojected into the earth model by the migration operator $\mathbf{L}^{\dagger}$. Least squares migration (LSM) is also known as linearized waveform inversion (Lailly, 1984; Tarantola, 1986, 1987; Jin et al., 1992; Lambaré et al., 1992) and is superior to standard migration by reducing migration artifacts caused by a poor acquisition geometry; it also can provide a spatial resolution that is more than twice (Nemeth et al., 1999; Yu et al., 2006) that of standard migration if the migration velocity is sufficiently accurate. Its main drawbacks are that its effectiveness is very sensitive to the accuracy of the migration velocity model, and it can be more than an order of magnitude more expensive than standard migration. (Recent developments [Dai and Schuster, 2009] in phase-encoded migration suggest a great reduction in the cost of LSM.) As an example, Figure 1d depicts the LSM image obtained from zero-offset (ZO) data, which is more accurate than the ZO standard migration image in Figure 1c. Figure 1e–f depict the RTM and LSM RTM images obtained from the prestack shot gathers.

Full waveform inversion

If the background model is updated after each iteration, then Equation A7 is known as nonlinear waveform inversion (Tarantola, 1987; Mora, 1987; Mora, 1989; Song et al., 1995; and many others); a common designation for waveform inversion is FWI or full waveform inversion. Unlike LSM or standard migration, Equation A7 for FWI iteratively updates the velocity and reflectivity models so that the final velocity image can be much improved in both resolution and accuracy. Its main drawbacks are a tendency to get stuck in local minima, it is computationally expensive compared to standard migration, and there might be more than one model that can explain the same data, i.e., a nonunique solution. For example, elastic isotropic modeling codes can sometimes generate predicted traces that adequately fit the observed data, but the estimated isotropic velocity model is inconsistent with the actual anisotropic rocks in the real earth.

Successful examples of waveform tomography images are shown in Figure 1g–h and were obtained from the same shot gathers used for migration. The final tomogram shows the velocity distribution that is almost identical to that of the actual model. One of the keys to success in waveform inversion is that a good starting model is often required for an accurate result. This starting model can be obtained by migration velocity analysis (Stork, 1992; Jousselin et al., 2009), reflection traveltime tomography (Langan et al., 1985; Bishop et al., 1985; Nolet, 1987; Zelt, 2002), or refraction tomography (Pratt and Goulty, 1991; Luo and Schuster, 1991; Min and Shin, 2006; Sheng et al., 2006).

A major challenge to the success of waveform inversion is the limited offset range between sources and receivers and the lack of low-frequency information in the recorded data (Sirgue and Pratt, 2004; Barnes et al., 2008; Boonyasiriwat et al., 2009; Kelly et al., 2009). Such deficiencies can prevent waveform inversion from reconstructing the low wave-number parts of the model, and sometimes prevent convergence to any type of

reasonable model. Remedies being explored include the possibility of recording data at much lower frequencies with more capable recording devices, and obtaining very wide-offset data. Other challenges address the validity of the acoustic approximation versus the reality that the recorded data are largely of elastic (Vigh et al., 2009) or viscoelastic nature (Causse et al., 1999; Askan, 2006). Elastic inversion of elastic seismograms (Mora, 1987; Mora, 1989; Zhou et al., 1997; Brossier et al., 2009) have been presented, but the acoustic approximation can still provide useful results. One approach to viscoelastic inversion is to invert for the acoustic velocity first, then follow this with inversion for attenuation parameters (Kamei and Pratt, 2008); another approach is to use a ray-based method to invert for the attenuation factor Q and then use this Q to correct for the attenuation in the data (Pratt et al., 2005; Sheng et al., 2007; Boonyasiriwat et al., 2009).

Wave equation traveltime inversion

If the data residual $\delta d(\mathbf{g}|\mathbf{s})$ is replaced by the traveltime residual $\delta\tau(\mathbf{g}|\mathbf{s})$ weighted by the recorded trace $d(\mathbf{g}|\mathbf{s})^{\text{obs.}}$, then this is known as wave equation traveltime (Luo and Schuster, 1991) tomography (WT); it is a variant of Rytov inversion (Woodward, 1992) and updates the velocity model by smearing weighted traveltime (not waveform) residuals over the associated wavepaths (Woodward, 1992; Luo, 1992). In the high-frequency limit, it reduces to ray-based traveltime tomography (RT). The advantage of WT over RT is that it does not require a high-frequency approximation and accounts for the band-limited nature of waves as they propagate through the earth. Shin et al. (2002, 2007) use a modified logarithm norm to invert for phase data, which bears a close relationship to the Rytov inversion method; and Effelsen (2009) compares phase inversion to traveltime tomography for inverting refraction events. Figure 2 illustrates how the above seismic imaging methods are related to one another.

The main disadvantage of WT is that it is at least an order of magnitude more expensive than RT because it requires a finite-difference solution of the wave equation rather than a ray-traced approximation. Earthquake seismologists often see wave equation traveltime as a major improvement in estimating the earth's velocity parameters from earthquake records (Dahlen et al., 2002; Marquering et al., 2002; Montelli et al., 2006; van der Hilst and de Hoop, 2006).

Migration velocity analysis

The previous imaging methods can be described as estimating the earth model by predicting synthetic data that *best matches the observed data in the data domain*. In contrast, migration velocity analysis (MVA) finds the velocity model that *best flattens or focuses the migration sections in the image domain*; here, the image domain is the migration cube in the (x, y, z) indices and some other index s such as shot index, receiver index, source-receiver offset, (Yilmaz and Chambers, 1984; Faye and Jeannot, 1986; Al-Yahya, 1989; Toldi, 1989; Stork, 1992; Lafond and Levander, 1993), common angle parameter (Xu et al., 1998), or subsurface offset coordinate (Rickett and Sava, 2002; Sava and Biondi, 2004; Robein, 2010).

To understand why we seek a velocity model that flattens events in the image domain, consider the 2D migration image $m(x, z, s_0)$ in Figure 3a obtained by migrating one shot gather (the source is at the surface with shot index s_0). If the migration velocity is correct, then all of the single-shot migration images should be similar in appearance. This means that all of the migrated reflection events should be flat in the common image gather (CIG) given by $m(x_{\text{CIG}}, z, s)$ for all values of s, z, and a fixed value of x_{CIG} (see Figure 3b). Note that if the migration velocity is accurate, the reflector boundary denoted by the dotted horizontal line will be flat for a common image gather. If the migration velocity is too slow, then the imaged reflector boundary will curve upward as illustrated by the curved dashed line, and if too fast, the imaged reflector will curve downward.

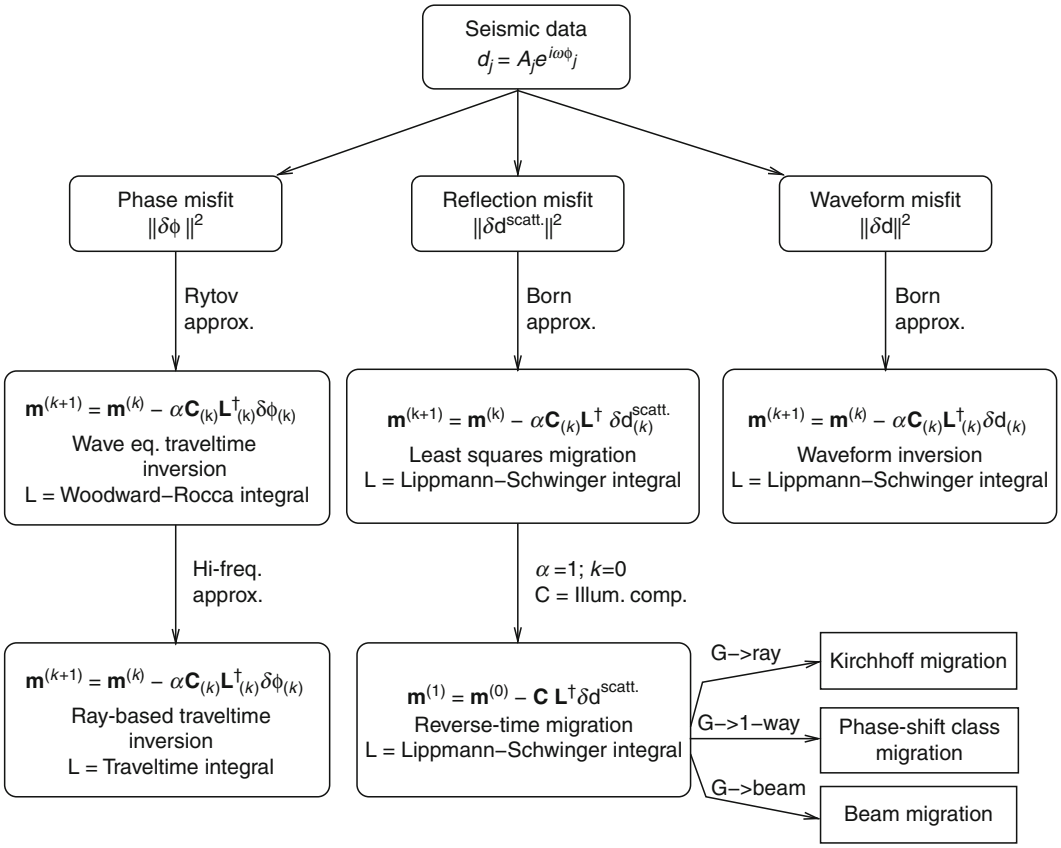
With MVA, the goal is to find the velocity model so that, ideally, the predicted migration image $\mathbf{m}_{\text{mig.}}^{\text{pred.}} = \mathbf{L}^{\dagger}\mathbf{L}\mathbf{M}$ best fits the actual migration image $\mathbf{m}_{\text{mig.}} = \mathbf{L}^{\dagger}\mathbf{d}$. The associated misfit function can be constructed so that it is similar to that of Equation 4, except that the norm of the *migration* residual

$$\varepsilon = \frac{1}{2}||\mathbf{m}_{\text{mig.}}^{\text{pred.}} - \mathbf{m}_{\text{mig.}}||_K^2 + \frac{\lambda^2}{2}||\mathbf{C}\mathbf{m}||_W^2 \tag{6}$$

is minimized rather than the *data* residual; in this case, $\mathbf{m}$ represents the velocity model. To find the velocity model, we can use the unregularized steepest descent equation:

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} - \overbrace{\alpha\mathbf{L}^{\dagger}\mathbf{L}\delta\mathbf{m}_{\text{mig.}}^{(k)}}^{\text{migration}-\text{modeling of migration residual}}, \tag{7}$$

where $\delta\mathbf{m}_{\text{mig.}}^{(k)} = \mathbf{m}_{\text{mig.}}^{\text{pred.}} - \mathbf{m}_{\text{mig.}}$ is the migration residual at the kth iteration. The importance of this formula is that it shows that the gradient term is computed by modeling the migration residual to get the *filtered* data residual $\mathbf{L}\delta\mathbf{m}_{\text{mig.}}^{(k)}$, and then the velocity model is updated by smearing this filtered (the filtered data residual is constructed by migrating and modeling the actual data residual; hence, it is a filtered version of the actual data residual) data residual into the model by the migration operation $\mathbf{L}^{\dagger}[\mathbf{L}\delta\mathbf{m}_{\text{mig.}}^{(k)}]$. These filtered residuals are smeared along wavepaths for each source-receiver pair of traces, and the migration and modeling operators are updated after each iteration. This leads to a moderate resolution of the velocity model because the filtered, not unfiltered, data are kinematically fitted. Moreover, MVA



Seismic Imaging, Overview, Figure 2 Three types of seismic imaging methods: phase inversion, least squares migration (LSM), and nonlinear waveform inversion. Note that LSM is a linearized inversion so that the modeling operator **L** does not get updated after each iteration. This compares to nonlinear waveform inversion which updates **L** after each iteration. The Lippmann–Schwinger-type and Woodward–Roca-type integrals are displayed in Table 1, and details about various migration methods are in the Seismic Migration section in the *Encyclopedia of Solid Earth Geophysics*. The traces that only contain scattered arrivals are obtained by muting the direct arrivals; and the ensemble of these traces is symbolized by the vector $\mathbf{d}^{scatt.}$. Instead of computing the Green's functions in the integral equations by a finite-difference solution to the wave equation, various approximations such as ray-based Green's functions (Kirchhoff migration), one-way wave equation approximations (phase-shift type migration), and Gaussian Beam (beam-like migration) are used for migration.

is largely a curve fitting exercise that is mostly insensitive to subtle amplitude variations in the migration traces; ignoring such information will blind MVA to subtle variations in the impedance distribution.

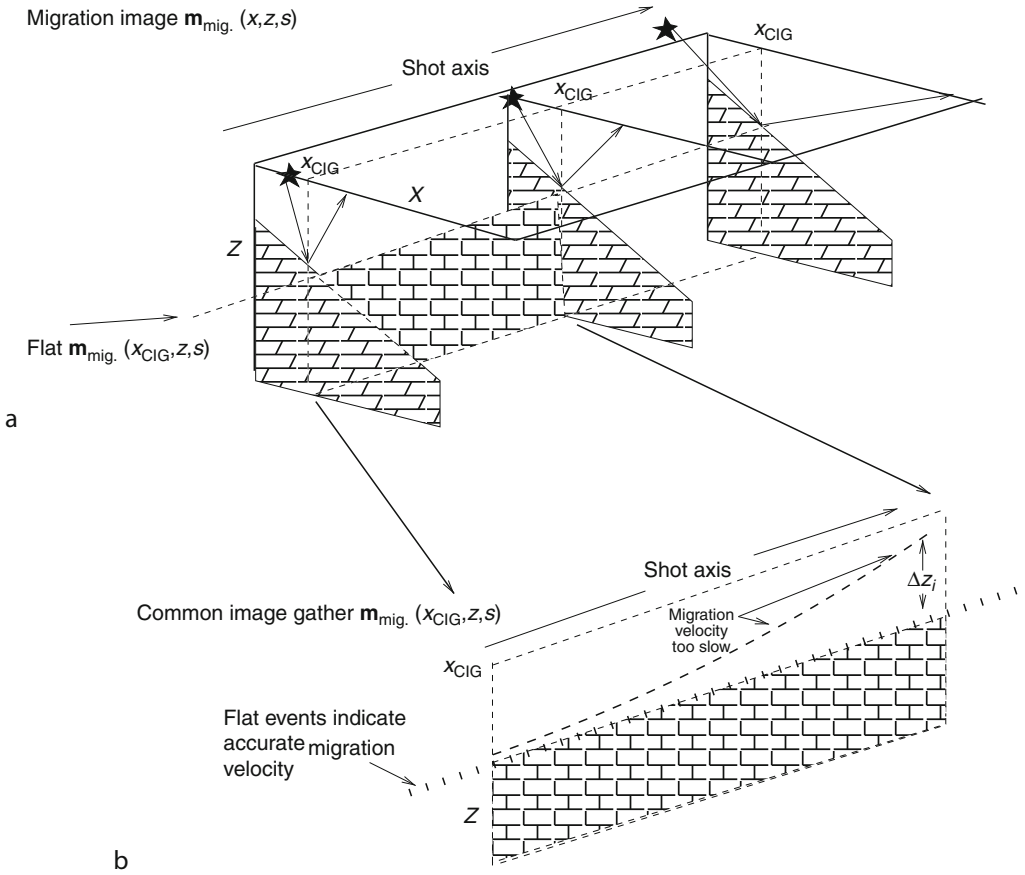
To reduce computational costs and increase robustness at the expense of reduced model resolution, MVA is sometimes implemented with the following steps:

1. Automatically pick the depth residual $\Delta z_i = z_i - z_i^{\text{ref.}}$ of a coherent CIG event in Figure 3b at the ith shot position; here, z_i is the picked depth of the targeted reflection at the x_{CIG} offset in the ith migrated shot gather. The depth $z_i^{\text{ref.}}$ of the reference event for that reflector is estimated from the near-offset trace in that CIG. A computer algorithm can window about the near-offset reflection of interest and use crosscorrelation with neighboring traces in the CIG to estimate the depth lag $z_i - z_i^{\text{ref.}}$ associated with the strongest correlation energy.
2. The misfit function is then defined as

$$\varepsilon = 1/2 \sum_i (\Delta z_i)^2 + \text{model smoothness constraints}, \tag{8}$$

for each CIG and the summation is over the shot index in Figure 3b. Sometimes the misfit function is the sum over all CIGs.
3. The velocity model is iteratively updated by a gradient optimization method until ε is minimized.

An example of the above procedure is shown in Figure 4, where the left column of images depict the (a) migration image obtained with an inaccurate velocity



Seismic Imaging, Overview, Figure 3 (**a**) Migration image $\mathbf{m}_{\text{mig.}}(x, z, s)$ cube in the model coordinates of x and z and the surface shot offset index s; the common image gather $\mathbf{m}_{\text{mig.}}(x_{CIG}, z, s)$ is computed by migrating the shot gathers and displaying the migration image at the fixed value of $x = x_{CIG}$ for all shot s and depth z values. (**b**) Common image gather $\mathbf{m}_{\text{mig.}}(x_{CIG}, z, s)$ for the common image point at x_{CIG}. The migration image should only be non-zero along the interface between the brick region and open regio shown above.

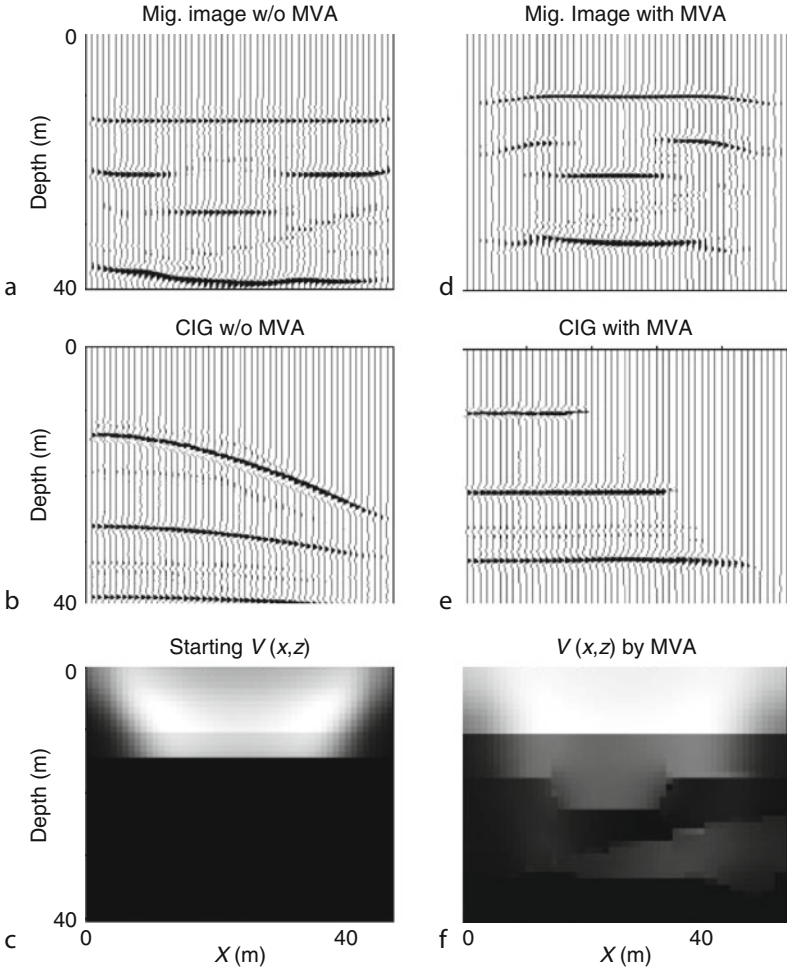
model, (b) the CIG with curved events, and (c) the migration velocity model. After seven iterations of MVA, the right column of figures is obtained. Note the flat events in the CIG, and the final velocity model is almost the same as the actual velocity model.

The Fréchet derivative $\frac{\partial z_i}{\partial c(\mathbf{X})}$ associated with the gradient of ε can be numerically computed by using a ray-tracing scheme (Stork, 1992; Chauris et al., 1998) to determine the change in depth z_i of a migrated reflection point with respect to a change in the velocity parameter at $\mathbf{x}$. Sometimes, the depth residual Δz_i for the ith ray is converted into a time residual $\Delta\tau_i$ and, similar to traveltime tomography, this *converted* time residual is smeared along the reflection ray to iteratively update the velocity model. For arbitrary reflector geometries in a homogeneous media, an analytic conversion formula was derived by Al-Yahya (1989), and a ray-tracing method was used by Stork (1992) and others (Robein, 2010).

Similar to waveform inversion, MVA seeks to predict the wiggly migration traces seen in the migration image $\mathbf{L}^{\dagger}\mathbf{d}$ and, therefore, can easily get stuck in the many local minima of ε. To avoid this problem, Symes and Carazone (1991) proposed the smoother differential semblance misfit function (Chauris et al., 1998; Shen et al., 2003):

$$\varepsilon = 1/2||\sum_{x,y}[\gamma\sum_{z}\sum_{h}\partial m(x,y,z,h)_{\text{mig.}}/\partial h^2]||^2 + \text{constraints}. \tag{9}$$

that rewards flat events (i.e., accurate velocity models) in the CIG domain. Here, γ is a normalization term that depends on (x, y). Empirical tests by Chauris et al. (1998) suggest that the semblance based ε is much smoother than that in Equation 6 and is nearly devoid of local minima in misfit function. Some links between MVA and FWI are established in Symes (2008), and MVA compared to several tomography methods is presented by Le Bégat et al. (2004). The subsurface offset domain can be exploited for MVA (Sava and Biondi, 2004) and extraction of scattering angle information



Seismic Imaging, Overview, Figure 4 (**a**) Migration image (Sun, 2001) obtained by prestack migration using the smooth homogeneous velocity model in (**c**) that is far from the true model approximated in (**f**); (**b**) common image gather in the shot offset index, where the curved events indicate an incorrect velocity model; (**c**) incorrect migration velocity model used to compute (**a**); (**d**) migration image obtained after seven iterations of MVA; (**e**) CIG after seven iterations of MVA; and (**f**) velocity model inverted by seven iterations of MVA. This result closely closely resembles the true velocity model.

(Rickett and Sava, 2002), and an excellent summary of MVA research is given in Robein (2010).

In summary, the exploration community heavily favors MVA over many other velocity estimation methods because it is robust, efficient, and the picking of depth residuals is easily automated in the migration image domain. (Traveltimes picked from traces are notoriously difficult to automate because waveforms often interfere with one another. In comparison, migration untangles these events and focuses them to their origin along the reflectors so that automatic picking is easier in the CIG. Sometimes semblance methods are used to find the best fit hyperbolic or parabolic curve to the data [Robein, 2010].) Its chief disadvantage is that an MVA image lacks the detailed resolution of waveform inversion, which suggests that MVA should be used to estimate the starting velocity models for waveform tomography.

Recent advances in seismic imaging

In the last 15 years, several breakthroughs have enabled practical waveform inversion of seismic data. One of these advances is the relentless increase in computational capabilities of cluster computers and GPU-like processors, and two others are multiscale inversion and phase-encoded waveform inversion.

Multiscale waveform inversion

One of the main difficulties with waveform inversion is that the misfit function is plagued by many local minima. If the starting velocity model is moderately far from the actual model (an erroneous velocity might be one where the modeled events arrive by more than a period after the actual arrivals), then the iterative gradient solution gets stuck there and never reaches the global minimum or actual model.

The partial cure to this local-minima problem is a multiscale approach (Bunks et al., 1995), where the initial iterations estimate a coarse-grid velocity model from low-frequency data. For a reasonable starting model, this often prevents getting stuck in local minima because the predicted low-frequency arrivals are more likely to arrive within a period of the arrivals in the low-pass filtered data. After a number of iterations, the velocity model is refined to a finer scale (the grid interval is halved) and intermediate frequency data are iteratively inverted to update the velocity model. After suitable reduction of the data misfit, the model grid is refined again and higher-frequency data are inverted until a satisfactory model is reconstructed. One of the first relevant demonstrations of multiscale FWI applied to real data in a complex environment was performed by Ravaut et al. (2004). Other results with both synthetic and field data (Sirgue and Pratt, 2004; Sirgue et al., 2007; Plessix, 2009; Vigh and Starr, 2007, 2008; Sirgue et al., 2010) convincingly demonstrate the effectiveness of this approach. As mentioned earlier, improved model reconstructions can be achieved if more accurate starting models are employed, lower frequency data are recorded, and wider offset data are acquired. The new challenges are to employ modeling and inversion codes that robustly take into account the effects of viscoelasticity and anisotropy in the data.

Phase-encoded multisource waveform inversion

A major difficulty, until recently, is the enormous computational expense of waveform inversion. Each shot gather of residuals must be migrated at each iteration, which for 3D problems can be too demanding even for the most powerful computers. To relieve this problem, Krebs et al. (2009), Virieux and Operto (2009), and Dai and Schuster (2009) proposed summing phase-encoded shot gathers into supergathers and migrating the supergathers at each iteration; a supergather is modeled with just one finite-difference simulation, where the computational cost is the same as that for one shot gather. This is similar to the phase-encoding RTM of Romero et al. (2000), except iterations are used to reduce the crosstalk noise in supergather migration. The result can be an enormous cost savings compared to conventional waveform inversion or migration.

Figure 1h shows an example of phase-encoding shot gathers, where each trace in a shot gather has the same random time shift but different shots have different time shifts. In this case, 12 shot gathers were time-shifted and blended together into one supergather; there were 192 traces per shot gather. A total of 16 nonoverlapping supergathers were input into the iterative waveform inversion code, where each supergather migration costs the same as the migration of one shot gather because 12 shots were excited at nearly the same time. Hence, the Figure 1h tomogram costs 1/12 that of the Figure 1g tomogram. For 3D inversion, the computational cost savings can be more than two orders of magnitude compared to conventional waveform inversion.

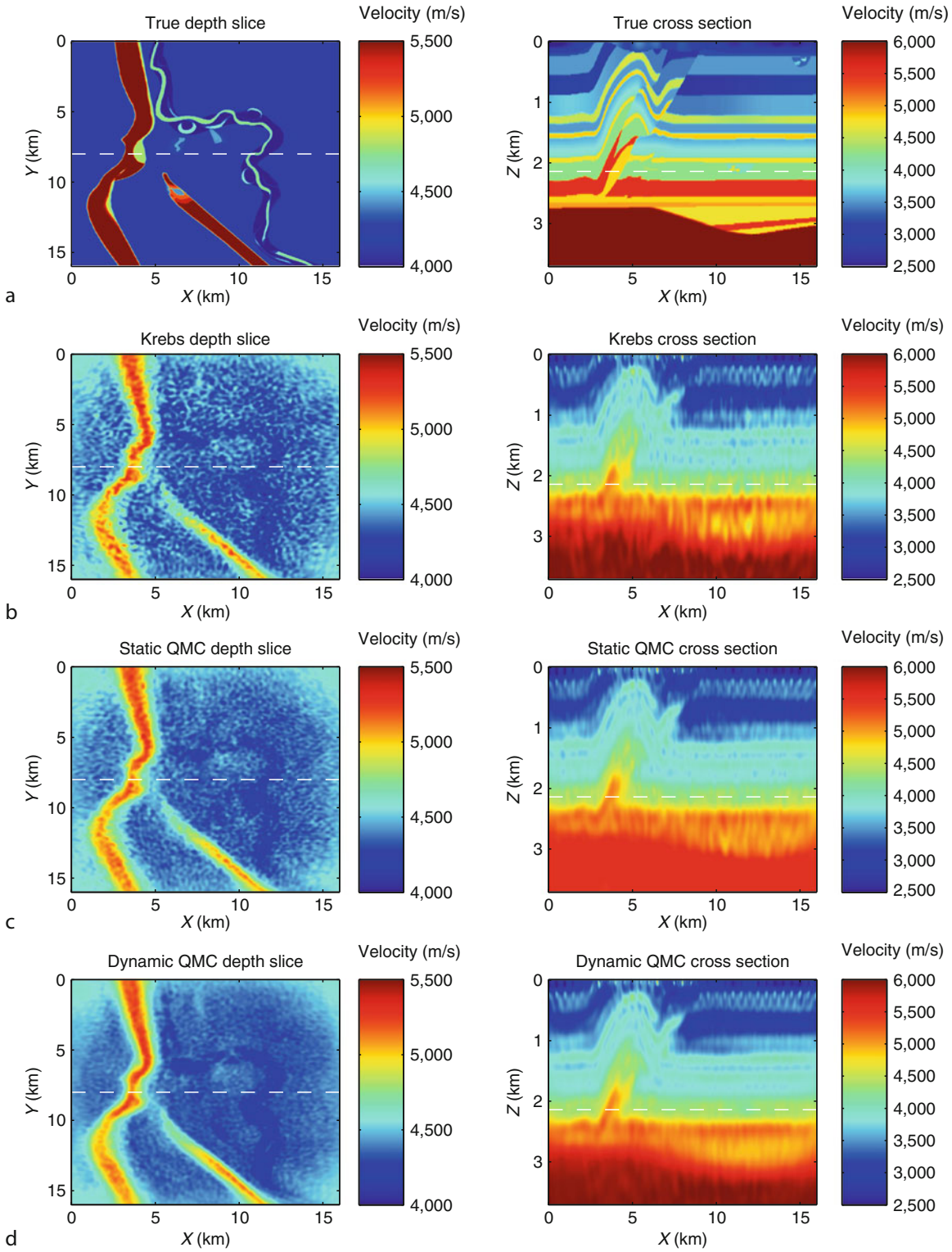
For a 3D example, Figure 5a shows results after waveform inversion of 3D synthetic data, where the model size is 800 × 800 × 186 grid points and the spatial sampling interval is 20 m. There are 1,089 sources evenly distributed along the surface with an interval of 500 m in the inline (X) and crossline (Y) directions. Multisource waveform inversion using the Krebs method (Krebs et al., 2009), static Quasi-Monte Carlo (QMC) phase encoding, and dynamic QMC phase encoding (Boonyasiriwat and Schuster, 2010) are applied to this data set.

Figure 5b–d show that dynamic QMC phase encoding provides a higher-quality tomogram than the other methods, yet the computational cost is two orders of magnitude less than that of conventional waveform inversion. The Krebs strategy is the most efficient because 1,089 shot gathers were blended into one supergather, compared to the QMC strategies that used 99 CSGs/supergather. Dynamic phase encoding changed the time shifts of each shot gather after each iteration, while the static strategy kept the phase encoding the same.

Current status and future of seismic imaging

Reverse time migration is becoming the preferred means for seismic imaging beneath complex regions such as salt bodies. Earthquake seismologists are now recognizing the benefits of migration imaging earthquake records for tectonic structures such as subduction zones (Bostock et al., 2001) or using wide-angle seismic experiments to image the crust (Brenders and Pratt, 2007) and mantle (Montelli et al., 2006). Least squares migration and waveform inversion are now being considered as viable upgrades to standard migration because of the tremendous speedup from phase-encoded multisource methodology. There is still a debate about whether waveform inversion should be computed in the frequency (there are additional advantages by formulating the problem in the Laplace transform domain [Shin and Ha, 2008]) or time domains (Vigh and Starr, 2007; Warner, 2008), but there is no debate that we eventually need to account for viscoelastic and anisotropic effects in the data.

3D waveform inversion of earthquake records for whole earth tomograms greater than 1 Hz is still too computationally demanding except at very low frequencies, and the same can be said for 3D exploration geophysics at frequencies above 50 Hz. Challenges still remain, especially in the critically important area of anisotropic RTM (Zhang and Zhang, 2009; Fowler et al., 2010) and waveform inversion; estimation of accurate anisotropic migration velocity models is an ongoing line of research. Earthquake seismologists are now testing the possibility of using earthquake records for inverting basin structures and velocity models so as to improve their simulation-based predictions of earthquake hazard. Passive seismic recordings and the use of interferometry to extract surface wave records (Shapiro and Campillo, 2004; Shapiro et al., 2005), followed by inversion for S-velocity tomograms are playing an increasingly important role in earthquake seismology.



Seismic Imaging, Overview, Figure 5 Depth slices at $z = 2.1$ km and cross-sections at $y = 8$ km of (**a**) true model, and inverted models using (**b**) the Krebs method (1089 CSGs/supergather), (**c**) static QMC method (99 CSGs/supergather), and (**d**) dynamic QMC method (99 CSGs/supergather) after 40 iterations. A greater number of CSGs/supergather require a greater number of iterations to get the same accuracy in the final tomogram; the dynamic encoding strategy is more effective than the static strategy. (Figures courtesy of C. Boonyasiriwat).

Bibliography

Al-Yahya, K., 1989. Velocity analysis by iterative profile migration. *Geophysics*, **54**, 718–729.

Aoki, N., and Schuster, G. T., 2009. Fast least-squares migration with a deblurring filter. *Geophysics*, **74**, WCA83–WCA93.

Askan, A., 2006. *Full Waveform Inversion for Seismic Velocity and Anelastic Losses in Heterogeneous Structures*. PhD dissertation, Carnegie Mellon University.

Barnes, C., Charara, M., and Tsuchiya, T., 2008. Feasibility study for an anisotropic full waveform inversion of cross-well seismic data. *Geophysical Prospecting*, **56**, 897–906.

Beydoun, W., and Mendes, M., 1989. Elastic ray-born L2-migration/inversion. *Geophysical Journal International*, **97**, 151–160.

Beylkin, G., Oristaglio, M., and Miller, D., 1985. Spatial resolution of migration algorithms. In Berkhout, A. J., Ridder, J., and van der Walls, L. F. (eds.), *Proceeding of the 14th International Symposium on Acoustical Imaging*. pp. 155–167.

Bishop, T., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., Spindler, D. A., and Wyld, H. W., 1985. Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, **50**, 903–923.

Bleistein, N., Cohen, J. K., and Stockwell, J. W., 2001. *Mathematics of multidimensional seismic inversion*. New York: Springer.

Boonyasiriwat, C., and Schuster, G. T., 2010. 3D multisource ful-waveform inversion using dynamic QMC phase encoding. In *Expanded Abstracts of SEG International Meeting*, p. 29.

Boonyasiriwat, C., Valasek, P., Routh, P., Cao, W., Schuster, G. T., and Macy, B., 2009. An efficient multiscale method for time-domain waveform tomography. *Geophysics*, **74**, WCC59–WCC68, doi:10.1190/1.3151869.

Bostock, M. G., Rondenay, S., and Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves: Theory for oblique incidence. *Journal of Geophysical Research*, **106**, 771–782.

Brenders, A. J., and Pratt, R. G., 2007. Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data. *Geophysical Journal International*, **168**, 152–170.

Brossier, R., Operto, S., and Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, **74**, 105–118.

Brossier, R., Operto, S., and Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, **75**, R37–R46.

Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G., 1995. Multiscale seismic waveform inversion. *Geophysics*, **60**, 1457–1473.

Causse, E., Mittet, R., and Ursin, B., 1999. Preconditioning for full-waveform inversion in viscoacoustic media. *Geophysics*, **64**, 130–145.

Chauris, H., Noble, M., and Podvin, P., 1998. Testing the behavior of differential semblance for velocity estimation. In *Expanded Abstracts of SEG International Meeting*, pp. i1305–1308.

Claerbout, J., 1992. *Earth Soundings Analysis: Processing versus Inversion*. Cambridge, MA: Blackwell Scientific.

Clinthorne, N., Pan, T., Chiao, P., Rogers, W., and Stamos, J., 1993. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Transactions on Medical Imaging*, **12**, 78–83.

Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A., 1990. Robust elastic non-linear waveform inversion: Application to real data. *Geophysics*, **55**, 527–538.

Dahlen, F., Hung, S., and Nolet, G., 2002. Frechet kernels for finite-frequency traveltimes I. Theory. *Geophysical Journal International*, **141**, 157–174.

Dai, W., and Schuster, G. T., 2009. Least-squares migration of simultaneous sources data with a deblurring filter. In *Expanded Abstracts of SEG International Meeting*, pp. 2990–2993.

Duquet, B., Marfurt, K., and Dellinger, J., 2000. Kirchhoff modeling, inversion for reflectivity, and subsurface illumination. *Geophysics*, **65**, 1195–1209.

Effelsen, K., 2009. A comparison of phase inversion and traveltime tomography for processing of near-surface refraction traveltimes. *Geophysics*, **74**, WCB11–WCB24.

Epanomeritakis, I., Akçelik, V., Ghattas, O., and Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. *Inverse Problems*, **24**, 975–987.

Erlangga, Y., and Hermann, F., 2009. Seismic waveform inversion with Gauss-Newton-Krylov method. In *Expanded Abstracts of SEG International Meeting*, pp. 2357–2361.

Etgen, J., Gray, S., and Zhang, Y., 2009. An overview of depth imaging in exploration geophysics. *Geophysics*, **74**, WCA5–WCA17.

Faye, J.-P., and Jeannot, J.-P., 1986. Prestack migration velocities from focusing depth analysis. In *Expanded Abstracts of SEG International Meeting*, pp. 438–440.

Fowler, P., Du, X., and Fletcher, R., 2010. Coupled equations for reverse time migration in transversely isotropic media. *Geophysics*, **75**, S11–S22.

Hill, N. R., 2001. Prestack Gaussian beam depth migration. *Geophysics*, **66**, 1240–1250.

Jin, S., Madariaga, R., Virieux, J., and Lambaré, G., 1992. Two-dimensional asymptotic iterative elastic inversion. *Geophysical Journal International*, **108**, 575–588.

Jousselin, P., Duquet, B., Audebert, F., and Sirgue, J., 2009. Bridging the gap between ray-based tomography and wave-equation migration image gathers. In *Expanded Abstracts of SEG International Meeting*, pp. 3979–3983.

Kamei, R., and Pratt, G., 2008. Waveform tomography strategies for imaging attenuation structure for cross-hole data. In *70th Conference and Technical Exhibition, EAGE Expanded Abstracts*, p. F019.

Kelly, S., Ramos-Martinez, J., and Tsimelzon, B., 2009. The effect of improved, low-frequency bandwidth in full-wave form inversion for velocity. In *Expanded Abstracts of SEG International Meeting*, pp. 3974–3977.

Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M. D., 2009. Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, **74**, WCC177–WCC188.

Lafond, C., and Levander, A., 1993. Migration moveout analysis and depth focusing. *Geophysics*, **58**, 91–100.

Lailly, P., 1984. Migration methods: partial but efficient solutions to the seismic inverse problem. In Santosa, F., Pao, Y. H., Symes, W., and Holland, C. H. (eds.), *Inverse Problems of Acoustic and Elastic Waves*. Philadelphia: SIAM.

Lambaré, Virieux, J., Madariaga, R., and Jin, S., 1992. Iterative asymptotic inversion in the acoustic approximation. *Geophysics*, **57**, 1138–1154.

Langan, R., Lerche, I., and Cutler, R. T., 1985. Tracing of rays through heterogeneous media: An accurate and efficient procedure. *Geophysics*, **50**, 1456–1465.

Le Bégat, S., Chauris, H., Devaux, V., Nguyen, S., and Noble, M., 2004. Velocity model estimation for depth imaging: Comparison of three tomography methods on a 2D real data set. *Geophysical Prospecting*, **52**, 427–438.

Luo, Y., 1992. Calculation of wavepaths for band-limited seismic waves. In *Expanded Abstracts of SEG International Meeting*, pp. 1509–1512.

Luo, Y., and Schuster, G. T., 1991. Wave-equation traveltime inversion. *Geophysics*, **56**, 645–653.

Marquering, H., Dahlen, F. A., and Nolet, G., 2002. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox. *Geophysical Journal International*, **137**, 805–815.

McMechan, G., 1983. Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, **31**, 413–420.

Min, D., and Shin, C., 2006. Refraction tomography using a waveform-inversion back-propagation technique. *Geophysics*, **71**(3), R21–R30.

Montelli, R., Nolet, G., Masters, G., Dahlen, F., and Hung, S. H., 2004. Global P and PP traveltime tomography: rays versus waves. *Geophysical Journal International*, **158**, 637–654.

Montelli, R., Nolet, G., and Dahlen, F., 2006. Comment on "Banana-doughnut kernels and mantle tomography" by van der Hilst and de Hoop. *Geophysical Journal International*, **167**, 1204–1210.

Mora, P., 1987. *Elastic Wavefield Inversion*. PhD dissertation, Stanford University, pp. 143.

Mora, P., 1989. Inversion=migration+tomography. *Geophysics*, **54**, 1575–1586.

Nemeth, T., Wu, C., and Schuster, G. T., 1999. Least-squares migration of incomplete reflection data. *Geophysics*, **64**, 208–221.

Nolet, G., 1987. *Seismic Tomography: with Applications in Global Seismology and Exploration*. Springer.

Nowack, R. L. W. P., Chen, U. K., and Dasgupta, S., 2007. Imaging offsets in the Moho: synthetic tests using Gaussian beams with teleseismic waves. *Pure and Applied Geophysics*, **164**, 1921–1936.

Operto, S., Virieux, J., Ribodetti, A., and Anderson, J., 2009. Finite-difference frequency-domain modeling of viscoelastic wave propagation in two-dimensional tilted transversely isotropic media. *Geophysics*, **74**, T75–T95.

Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**, 495–503.

Plessix, R. E., 2009. 3D frequency-domain full-waveform inversion with an iterative solver: Geophysics. *Geophysics*, **74**, WCC149–WCC157.

Pratt, G., 1999. Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model. *Geophysics*, **64**, 888–901.

Pratt, R. G., and Goulty, N. R., 1991. Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data. *Geophysics*, **56**, 208–224.

Pratt, R. G., Shin, C., and Hicks, G. J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, **133**, 341–362.

Pratt, R. G., Hou, F., Bauer, K., and Weber, M., 2005. Waveform tomography images of velocity and inelastic attenuation from the Mallik 2002 crosshole seismic surveys. In Dallimore, S. R., and Collett, T. S. (eds.), *Scientific Results from the Mallik 2002 Gas Hydrate Production Research Well Program*. Mackenzie Delta, Northwest Territories, Canada: Geological Survey of Canada.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and dell'Aversana, P., 2004. Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International*, **159**, 1032–1056.

Rickett, J., and Sava, P., 2002. Offset and angle-domain common-image gathers for shot-profile migration. *Geophysics*, **67**, 883–889.

Robein, E., 2010. *Seismic Imaging*. EAGE Publications.

Romero, L., Ghiglia, D., Ober, C., and Morton, S., 2000. Phase encoding of shot records in prestack migration. *Geophysics*, **65**, 426–436.

Ryberg, T., and Weber, M., 2000. Receiver function arrays: a reflection seismic approach:. *Geophysical Journal International*, **141**, 1–11.

Sava, P., and Biondi, B., 2004. Wave-equation migration velocity analysis-1: theory. *Geophysical Prospecting*, **52**, 593–606.

Shapiro, N., and Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, **31**, L07614, doi:10.1029/2004GL019491.

Shapiro, N., Campillo, M., Stehly, L., and Ritzwoller, M., 2005. High-resolution surface-wave tomography from ambient seismic noise. *Science*, **307**, 1615–1618.

Sheehan, A. F., Shearer, P. M., Gilbert, H. J., and Dueker, K. G., 2000. Seismic migration processing of P-SV converted phases for mantle discontinuity structure beneath the Snake River Plain, Western United States. *Journal of Geophysical Research*, **105**, 055–065.

Shen, P., Symes, W., and Stolk, C., 2003. Differential semblance velocity analysis by wave equation migration. In *Expanded Abstracts of SEG International Meeting*, pp. 2135–2139.

Sheng, J., Leeds, A., Buddensiek, M., and Schuster, G. T., 2006. Early arrival waveform tomography on near-surface refraction data. *Geophysics*, **71**(4), U47–U57.

Shin, C., and Ha, W., 2008. A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains. *Geophysics*, **73**, VE119–VE133.

Shin, C., Min, D.-J., Marfurt, K. J., Lim, H. Y., Yang, D., Cha, Y., Ko, S., Yoon, K., Ha, T., and Hong, S., 2002. Traveltime and amplitude calculations using the damped wave solution. *Geophysics*, **67**, 1637–1647.

Shin, C., Pyun, S., and Bednar, B., 2007. Comparison of waveform inversion, part 1: conventional wavefield vs logarithmic wavefield. *Geophysical Prospecting*, **55**, 449–464.

Sirgue, L., and Pratt, R. G., 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, **69**, 231–248.

Sirgue, L., Etgen, J., and Albertin, U., 2007. 3D full-waveform inversion: wide-versus narrow-azimuth acquisitions. In *Expanded Abstracts of SEG International Meeting*, pp. 1760–1764.

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H., 2010. Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, **28**, 65–70.

Song, Z., Williamson, P., and Pratt, G., 1995. Frequency-domain acoustic-wave modeling and inversion of crosshole data, Part 2: Inversion method, synthetic experiments and real-data results. *Geophysics*, **60**, 786–809.

Stolt, R., and Benson, A., 1986. Seismic migration: theory and practice. In *Handbook of Geophysical Exploration*. London, UK: Geophysical Press, Vol. 5.

Stork, C., 1992. Reflection tomography in the postmigrated domain. *Geophysics*, **57**, 680–682.

Sun, H., 2001. *Wavepath Migration for Depth Imaging and Velocity Analysis*. PhD dissertation, University of Utah.

Symes, W., 2008. Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, **56**, 765–790.

Symes, W., and Carazone, J., 1991. Velocity inversion by differential semblance optimization. *Geophysics*, **56**, 654–663.

Tarantola, A., 1986. Linearized inversion of seismic reflection data. *Geophysical Prospecting*, **32**, 998–1015.

Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Amsterdam: Elsevier Science.

Thierry, P., Operto, S., and Lambaré, G., 1999. Fast 2D ray-Born inversion/migration in complex media. *Geophysics*, **64**, 162–181.

Toldi, J., 1989. Velocity analysis without picking. *Geophysics*, **54**, 191–199.
van der Hilst, R., and de Hoop, M., 2006. Reply to comment by R. Montelli, G. Nolet, and F.A. Dahlen on "Banana-doughnut kernels and mantle tomography". *Geophysical Journal International*, **167**, 1211–1214.
Vigh, D., and Starr, E. W., 2007. Comparisons for waveform inversion, time domain or frequency domain? In *Expanded Abstracts of SEG International Meeting*, pp. 1890–1894.
Vigh, D., and Starr, E. W., 2008. 3D prestack plane-wave full-waveform inversion. *Geophysics*, **73**, 135–144.
Vigh, D., Starr, E. W., and Elapavuluri, P., 2009. Acoustic waveform inversion vs. elastic data. In *Expanded Abstracts of SEG International Meeting*, pp. 2298–2301.
Virieux, J., and Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**, WCC1–WCC26.
Warner, 2008. 3D wavefield tomography: synthetic and field data examples. In *Expanded Abstracts of SEG International Meeting*, pp. 3330–3334.
Whitmore, N. D., 1983. Iterative depth migration by backward time propagation. In *Expanded Abstracts of SEG International Meeting*, pp. 827–830.
Williamson, P., 1991. A guide to the limits of resolution imposed by scattering in ray tomography. *Geophysics*, **56**, 202–207.
Woodward, M. J., 1992. Wave-equation tomography. *Geophysics*, **57**, 1526.
Xu, W., and Xie, X., 2009. How serious is the nonlinear effect on traveltime delays predicted by sensitivity kernels. In *Expanded Abstracts of SEG International Meeting*, pp. 4049–4053.
Xu, S., Chauris, H., Lambar, G., and Noble, M., 1998. Common angle image gather: a strategy for imaging complex media. In *Expanded Abstracts of SEG International Meeting*, pp. 1538–1541.
Yilmaz, O., and Chambers, R., 1984. Migration velocity analysis by wave-field extrapolation. *Geophysics*, **49**, 1664–1674.
Yu, J., Hu, J., Schuster, G., and Estill, R., 2006. Prestack migration deconvolution. *Geophysics*, **71**, S53, doi:10.1190/1.2187783.
Zelt, C., 2002. Modelling strategies and model assessment for wide-angle seismic traveltime data. *Geophysical Journal International*, **139**, 183–204.
Zhan, G., and Schuster, G. T., (in press). Multisource phase-encoded waveform inversion. *Journal of Exploration Geophysics*.
Zhang, Y., and Zhang, H., 2009. A stable TTI reverse time migration and its implementation. In *Expanded Abstracts of SEG International Meeting*, pp. 2794–2798.
Zhou, C., Schuster, G. T., Hassanzadeh, S., and Harris, J. M., 1997. Elastic wave-equation traveltime and waveform inversion of crosshole seismic data. *Geophysics*, **62**, 853–868.

Cross-references

Body Waves
Free Oscillations of the Earth
Inverse Theory, Global Optimization
Inverse Theory, Linear
Inverse Theory, Monte Carlo Method
Numerical Methods, Finite Difference
Seismic Tomography
Seismic, Ambient Noise Correlation
Seismic, Migration
Seismic, Waveform Modeling and Tomography
Single and Multichannel Seismics
Traveltime Tomography Using Controlled-Source Seismic Data
Vertical Seismic Profiling

SEISMIC INSTRUMENTATION

Duncan Carr Agnew
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

Synonyms

Seismometry

Definition

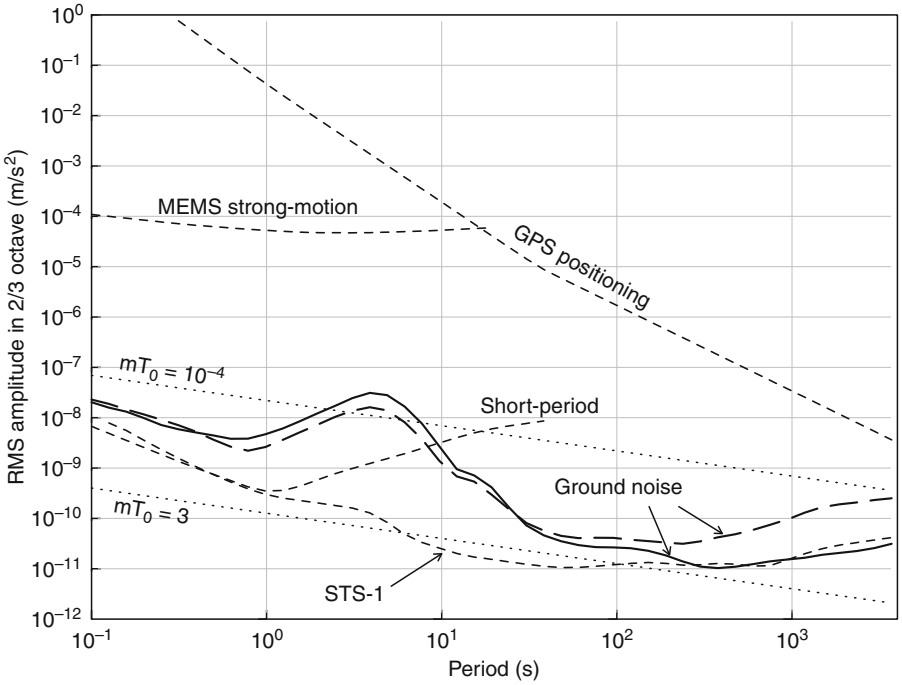
Datalogger. Device for recording electrical signal from a seismometer, usually in digital form and with accurate time information.
Geophone. Another name for seismometer, used in geophysical exploration.
Seismometer. Device for providing a record of ground motion, usually by converting it into an electrical signal.
Strainmeter. Device for continuously measuring the deformation of the Earth, either as extension along a line, or volume change.

Introduction

Seismic instruments give a time record of ground motion caused by elastic waves. The first instruments were built at the end of the nineteenth century; subsequent developments have evolved towards wider coverage of both the frequency and amplitude of the waves that could be recorded. Over most of the twentieth century much of this evolution, and of the diversity of instrument types, was related not to how ground motion was sensed but to how it was recorded; the last 3 decades have seen such rapid development in digital technology, especially in data storage, that recording techniques unique to seismology are no longer needed. Modern seismic systems all consist of a seismometer for converting some aspect of ground motion into an electrical signal, and a datalogger for recording this signal. In exploration geophysics, the name geophone is a common synonym for seismometer; now that data recording is clearly separated from sensing ground motion, the old term seismograph could probably be abandoned. Since dataloggers are not unique to seismology, this article discusses only the general requirements for them, focusing instead on the designs of seismic sensors, or seismometry. Three recent reviews cover seismometry in more detail: Wielandt (2002), Havskov and Alguacil (2004), and Bormann (2009).

Requirements for instruments

The frequencies to be measured by seismometers range from 0.31 mHz (the slowest free oscillation) up to 1 kHz (in near-surface geophysics): roughly 6 orders of magnitude. Figure 1 shows the amplitudes that can occur over most of this frequency range, from the background noise level at quiet sites (Berger et al., 2004) to the large signals that have been observed at moderate distances from very



Seismic Instrumentation, Figure 1 Seismic and instrument noise levels. The solid and long dashed lines show the lowest levels of vertical and horizontal ground motion observed by the stations of the Global Seismographic Network. The dotted lines show the seismometer Brownian noise limit for two different values of the product of the sensor mass and free period. The dashed lines show the noise levels for a high-quality short-period system using a moving-coil transducer (Rodgers, 1994), for a MEMS sensor used for measuring strong motion (Evans et al., 2005), for a sensitive broadband system (Ringler and Hutt, 2009), and for GPS used at seismic frequencies.

large earthquakes, to the even larger, and damaging, strong ground motions close to an earthquake. The large signals are infrequent, but also of considerable scientific importance – but so are very small signals, which can be seen only if the seismometer has lower noise than the ground does.

Since digital dataloggers can record over a much wider range of amplitudes and frequencies than the older analog recorders could, modern seismic systems can be characterized as broadband (covering a wide band of frequencies) and high dynamic range (a wide range of amplitudes). But the range from noise to the largest signal covers over 10 orders of magnitude, too much for any single instrument and datalogger. Since most seismometer designs require practical tradeoffs between low-frequency capability and other performance goals, seismometers are still classified by what frequencies and amplitudes they cover best. Systems for recording the largest signals without distortion are called *strong-motion* seismometers, and have usually been designed to meet the needs of earthquake engineers rather than seismologists, though digital technology is blurring this distinction.

Inertial seismometers: basic principles

Almost all seismometers measure motion of the ground using the inertia of an internal mass, and so are called *inertial* sensors. This mass, of amount m, is subject to five forces: (1) constraints that restrict the mass motion to be in a particular direction (denoted by a unit vector $\mathbf{e}$) making the system one with a single degree of freedom; (2) the gravitational force vector, $\mathbf{g}m$, along that direction, namely, $\mathbf{e} \cdot \mathbf{g}m$; (3) an elastic restoring force (from a spring), $-k(x - x_0)$, where k is the spring constant and x the displacement (in the direction $\mathbf{e}$) away from an equilibrium position x_0; (4) a viscous damping force proportional to velocity $-d\dot{x}$; (5) additional forces F_b that may be applied by a feedback system (discussed below).

The displacement of the mass relative to an inertial frame is $x + \mathbf{e} \cdot \mathbf{u}$, where x is the motion of the mass relative to the Earth, which is what we can measure, and $\mathbf{u}$ is the (vector) motion of the Earth relative to the inertial frame, which is that of the Earth in the absence of seismic waves (we can neglect non-inertial effects from the Earth's rotation). Combining forces (2), (3), and (4), and applying the momentum equation, the acceleration of the mass is given by

$$m(\ddot{x} + \mathbf{e} \cdot \ddot{\mathbf{u}}) = \mathbf{e} \cdot \mathbf{g}m - k(x - x_0) - d\dot{x}$$

where we allow the possibility of variations in both $\mathbf{e}$ and $\mathbf{g}$. With no motion, we assume that the mass position is x_0, so that $\mathbf{e}_0 \cdot \mathbf{g}_0 m = -kx_0$. Then the above equation

becomes, for small variations in **e** and **g**, and after some rearrangement of terms,

$$\frac{k}{m}x + \frac{d}{m}\dot{x} + \ddot{x} = -\mathbf{e}_0 \cdot \ddot{\mathbf{u}} - \mathbf{e}_0 \cdot (\mathbf{g} - \mathbf{g}_0) - (\mathbf{e} - \mathbf{e}_0) \cdot \mathbf{g}_0$$

more usually written as

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 x = -\mathbf{e}_0 \cdot \ddot{\mathbf{u}} - \mathbf{e}_0 \cdot (\mathbf{g} - \mathbf{g}_0) - (\mathbf{e} - \mathbf{e}_0) \cdot \mathbf{g}_0 \quad (1)$$

where $\omega_0 = \sqrt{k/M}$ is the natural frequency of the seismometer ($T_0 = 2\pi/\omega_0$ is called the free period); $\gamma = d/2M$ is the damping constant. These names describe the behavior of the seismometer if the right-hand side of the equation is zero and $x = x_0$ at $t = 0$; the subsequent motion of the seismometer mass is then

$$x(t) = x_0 cos\left(\frac{2\pi t}{T_0}\right) e^{-\gamma t},$$

which is a decaying sinusoid with period T; the damping γ needs to be large enough to avoid ringing (a sign of resonance effects), and is usually set to about 0.8.

The right-hand side of Equation 1 shows that an inertial sensor responds to three inputs:

1. The acceleration of the ground, $\ddot{\mathbf{u}}$; this is why inertial sensors are often called accelerometers.
2. Changes in the gravitational vector **g** along the direction of sensitivity; if this is the dominant signal, the sensor is usually referred to as a gravimeter (see *Gravimeters*).
3. Changes in the direction of sensitivity relative to the direction of gravity. If $\mathbf{e}_0$ and $\mathbf{g}_0$ are parallel (the mass moves vertically), then to first order $\mathbf{e} - \mathbf{e}_0$ is perpendicular to $\mathbf{g}_0$, and this contribution is zero. However, if $\mathbf{e}_0$ and $\mathbf{g}_0$ are perpendicular (the mass moves horizontally), then to first order $\mathbf{e} - \mathbf{e}_0$ is parallel to $\mathbf{g}_0$, and this contribution can be significant; if this is the dominant signal, the sensor would be referred to as a tiltmeter, since changes in **e** reflect tilting of the sensor (or the ground it is attached to).

Much confusion has been created by the use of these different terms for what is the same kind of sensor, when the difference actually refers to the type of signal being measured.

Ignoring the tilt and gravity terms, suppose that the ground displacement is purely sinusoidal, $u(t) = Ue^{i\omega t}$ (U being complex, and the real part being taken). Then the mass motion will be $Xe^{i\omega t}$, where

$$\frac{X}{U} = \frac{\omega^2}{\omega_0^2 - \gamma i\omega + \omega^2} \quad (2)$$

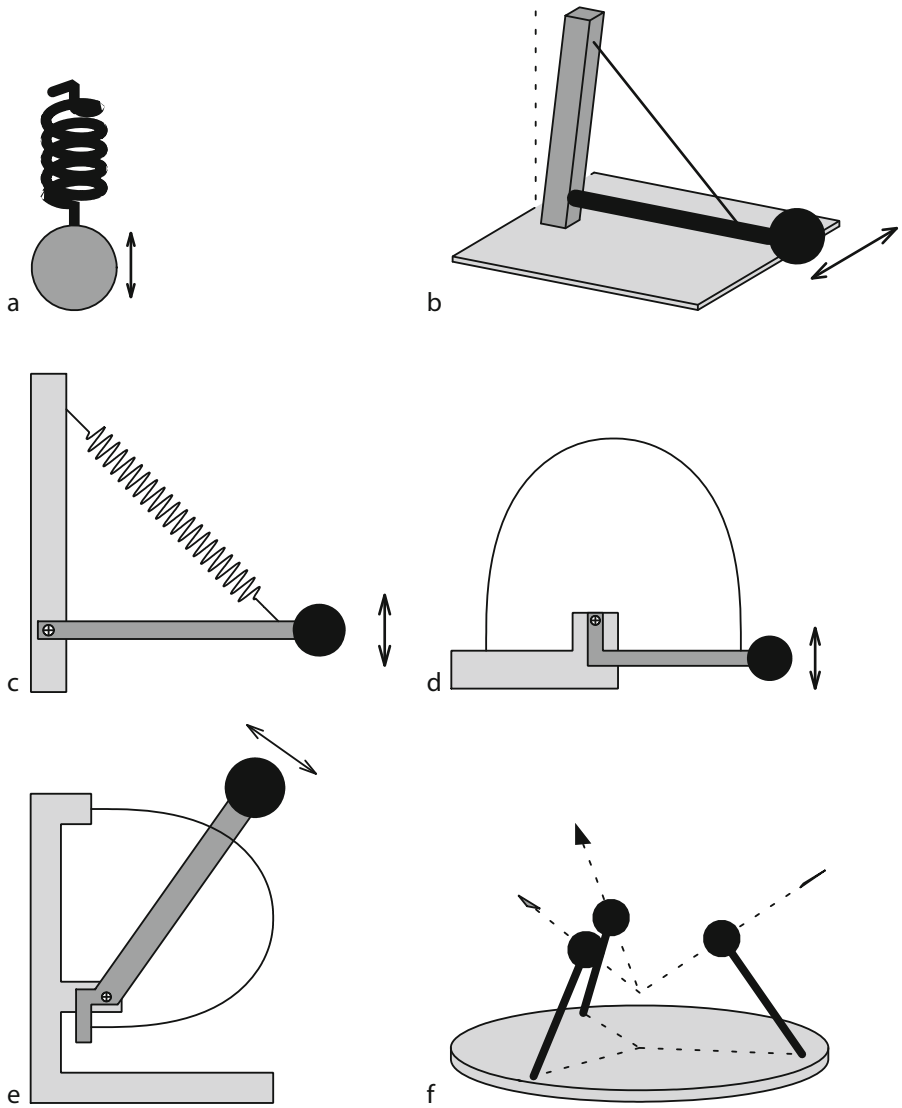
is the frequency response of the seismometer. For ground motion at frequencies much higher than the natural frequency ω_0, $X \approx U$: the mass motion looks like ground displacement. For ground motion at much lower frequencies, $X \approx U\omega^2/\omega_0^2$: the mass motion looks like ground acceleration, scaled by the inverse of the natural frequency of the seismometer.

The minimum noise level of any inertial sensor is set by the thermal (Brownian) motion of the mass in equilibrium with its surroundings. This noise level can be expressed as an equivalent ground acceleration, with a flat power spectral density of $8\omega_0\gamma k_B\theta/m$, where k_B is Boltzmann's constant and θ is the temperature (Aki and Richards, 2002). For a temperature of 310 K and $\gamma = 0.8$, this expression is $1.72 \times 10^{-19}/(mT_0)m^4s^{-3}$, with m in kilograms and T_0 in seconds. The dotted lines in Figure 1 show noise levels for two cases that cover the range of designs: $mT_0 = 10^{-4}$ (1 g mass and free period of 0.1 s), which cannot resolve ground noise at the longer periods, and $mT_0 = 3$ (300 g and 10 s), which shows that a moderately large mass and long period are necessary and sufficient for this thermal noise limit to fall below ground noise at all frequencies.

Mechanical design

In an inertial sensor, the mass needs to be constrained to move in one direction only, with a restoring force that is both exactly linear in the mass displacement and of whatever amount needed for the sensor to have the desired free period. The free period enters into the design partly because of Equation 2: the longer the period, the greater the displacement of the mass for a given acceleration at periods longer than the free period; that is to say, the higher the long-period sensitivity, and the lower the noise. Sensors designed to measure only high frequencies, such as the geophones used in geophysical exploration, can have a short free period, which means that the springs used can be relatively stiff (large k) and rugged. Seismometers for high frequencies use elastic elements both to constrain mass motion to a straight line and to provide a linear restoring force (Figure 2a). The latest innovation for short-period instruments is the MEMS (Micro-Electronic Mechanical Systems) sensor, in which the mass-spring system is machined from monolithic silicon and packaged with integral electronics; these are mass-produced for non-seismological applications and so are relatively inexpensive. While not as sensitive as traditional seismometers, they are already useful for measuring strong ground motions.

The high sensitivity needed for scientific purposes makes the mechanical design more challenging. If the mass moves vertically, it must be stably suspended against gravity while the restoring force is also kept low. In most long-period seismometers, the mass is constrained by hinges and stops to move along a small part of the circumference of a circle, which for most applications is an adequate approximation to a straight line (departures from this cause cross-coupling [Rodgers, 1968; LaCoste, 1967]). For sensing horizontal motion, the circle is slightly tilted from a horizontal plane, in which case the hinged



Seismic Instrumentation, Figure 2 Mechanical designs for seismometers, shown in cartoon form. (**a**) is a simple mass on a spring. (**b**) is the "garden-gate" design used for long-period horizontal instruments; note that the support post is not vertical. (**c**) shows the geometry of a Lacoste suspension; the restoring force of the spring must be equal to its physical length. (**d**) shows a vertical sensor that uses a leaf spring (elastica) to supply the restoring force. (**e**) and (**f**) show the design of a single sensor in a triaxial seismometer, and how three such sensors are arranged to record orthogonal motions.

mass is a horizontal pendulum (also called a garden-gate suspension, shown in Figure 2b). The restoring force then comes from gravity and is easily adjusted by tilting the instrument.

Long-period sensors that measure in other directions use a spring arranged to give a weak, but still linear, restoring force. One common design is the LaCoste suspension (Figure 2c), which uses a *zero-length spring*: a helical spring that exerts a force exactly proportional to its physical length. The geometry of the spring and hinged mass is chosen to make the restoring force nearly independent of mass position, giving a very long period. An alternative method uses a single sheet of elastic material, bent to form a leaf spring (Figure 2d), to create a long-period system with adequate linearity over the actual range of motion.

The traditional arrangement of seismometers is an orthogonal pair for horizontal motion, and a single sensor of vertical motion; if only one sensor is used, it is usually a vertical to give the best signal-to-noise ratio for P waves. Many systems now use a "triaxial" arrangement, with three identical systems measuring motion at an angle of 54.73° to the vertical, and at azimuths 120° apart; this also gives three orthogonal directions, namely those of the edges of a cube balanced on a corner (Figure 2e and f).

Any seismometer requires careful design to ensure that the only vibration that takes place is the mass motion, or at

least that other modes of vibration (called *parasitic resonances*) will not produce any signals if excited by ground motion. Long-period seismometers need special materials that do not creep much under load and whose dimensions and elastic constants are insensitive to temperature changes; even then, isolation and active environmental control may be needed to reduce the effects of changes in air temperature, air pressure (which causes varying buoyancy of the mass), and magnetic fields (which affect spring materials).

Electronic design

Sensing the mass motion in a seismometer is now almost always done electronically, in one of two ways (Agnew, 1986). In the first, electrodynamic method, permanent magnets and coils of wire are arranged so that the motion of the sensor produces an induced voltage $V = G\dot{x}$, where G is the generator constant; this voltage is then amplified and recorded. This method is simple and rugged, and provides damping if some of the induced current flows through a resistor. However, because the voltage produced depends on the velocity of the mass, and because the noise in the amplifiers rises with low frequency, electrodynamic sensing cannot resolve ground noise at periods longer than about 10 s.

The second method is active sensing of the mass motion. An oscillating voltage is input to a variable capacitor or inductor attached to the mass, which produces an output voltage with the same frequency and amplitude proportional to x; this output voltage is then demodulated to produce $V = Sx$, where S is the sensitivity. Up to the point of demodulation, the signal is at a frequency (several kHz) at which electronics noise is relatively low, so such displacement sensors can be made sensitive enough to measure motions less than an atomic diameter.

This output voltage can be recorded directly, but more often is used in a feedback system: the integral of this voltage (over some frequency band) is used to apply an additional force to the mass, usually with a coil and magnet, acting against the mass displacement. Because this force acts against the apparent force of ground acceleration, this is called a force-balance accelerometer. Feedback does not decrease noise, but does have three other merits. First, it can be used to vary the seismometer response much more than mechanical modifications can; for example, a mechanical system with a free period longer than 10 s is difficult to make, but this is easy to create electronically; and the response of the electronic system will be much more stable over time. Second, because the feedback force reduces the mass motion, the mechanical system needs to be linear over a smaller range. Finally, the calibration depends on the relation between voltage and feedback force, and this is very stable over time.

Calibration

Equation 2 is an example of the transfer function of a seismometer: the relation, as a function of frequency, between the ground motion and the output. In that example, the output was assumed to be mass motion x; a more realistic example would include a displacement sensor that generates a voltage $V = Sx$ followed by a datalogger that digitizes the voltage and produces a number $N = AV$. The total transfer function of the system, from input Earth displacement to digital output, would then be

$$\frac{AS\omega^2}{\omega_0^2 - \gamma i\omega + \omega^2}. \tag{3}$$

Seismometer calibration consists of determining all the parameters in expressions like this one. Some components, such as the datalogger, are calibrated by putting a known signal in and recording the output. For the seismic sensor, this requires producing a known ground motion, which can be done with a shake table before the seismometer is installed. After installation, a seismometer calibration can be checked by putting a signal into the seismometer electronics that will apply force to the mass without otherwise disturbing the output.

Calibration information is usually provided in a form like Equation 3: a leading constant, and the (complex) polynomials that give the response as a function of a frequency. Often the polynomials are specified by giving their roots in the complex plane, referred to as poles and zeros. The frequency response can usually be found quite accurately, though the absolute response to ground motion is usually difficult to estimate to better than 1%. Sometimes mundane aspects of the calibration, such as the actual direction of measurement, can be in error by significant amounts.

Installation

An ideal seismometer installation should maximize signals of interest whether distant earthquakes or the shaking of a building, while minimizing noise that might hide these signals. Any actual installation will be a compromise between this ideal and such practical matters as access and cost. For earthquake signals, putting an instrument deeper below the Earth's surface will give better results, but if this requires a deep drill-hole, a shallower installation may be all that can be done. An extreme example of low-quality installations being accepted for reasons of cost is ocean-bottom seismometers, for which the standard installation is the seismometer sitting on the bottom, sometimes on very soft sediments, however it happens to land after being dropped off a ship.

On land, the shallowest installations are for temporary instruments used in geophysical exploration, with seismometers planted on the surface with an attached spike to hold them in place. More permanent installations range from instruments set on surface rock exposures or in shallow holes in soil, to purpose-built deep vaults, caves (if available), and drill-holes. Deeper installations reduce the noise from wind and the range of temperature variation, and even a meter of soil will attenuate daily fluctuations. It is advisable, though not always possible, for instruments to be installed away from noise sources such as roads, machinery, rivers, and trees; in an urban setting borehole

installations at depths of a few hundred meters may be the only way to reduce the noise to reasonable levels.

Displacement seismometry using satellites

Because the mass motion is proportional to acceleration for motions with periods longer than the seismometer free period, data from inertial seismometers has to be integrated twice to find long-period displacements. This integration can introduce significant error, especially if the response is slightly nonlinear or the sensor tilts.

Direct sensing of displacement can now be done using repeated distance measurements between a ground sensor and Earth satellites, since the satellites define an inertial reference frame unaffected by ground motion. Distances between satellites of the Global Positioning System (GPS) and a receiver on the ground can be measured to within less than a millimeter precision, even over short times and in the presence of large accelerations; the actual ground displacement can be found after some processing (Bilich et al., 2008). The accuracy of GPS distance measurements is lower than their precision because of propagation effects, including interference from signals reflected from nearby objects (known as multipath), but it has proved possible to determine horizontal ground displacements to within a few millimeters over time spans up to several minutes and at rates up to 10 Hz. This sensitivity is too low to measure most earthquake signals, but can provide high-quality displacement data for ground motions near large earthquakes; a combination of inertial and GPS sensors is the optimal way to measure strong motion.

Deformation seismometers

A final class of seismometers measures, not the displacement at a point, but quantities related to the displacement gradient. For a vector displacement $\mathbf{u}$, the gradient $\nabla\mathbf{u}$ is a tensor, which can be decomposed into two parts: a symmetric part, the strain tensor, which describes local deformation; and an antisymmetric part, which is equivalent to a vector that describes local rotation, tilting about horizontal axes and rotating about the vertical axis.

Two types of instruments, both known as strainmeters, can measure components of the strain tensor (Agnew, 1986). Extensometers measure relative displacement along a line between two end points; this displacement divided by the baseline (the distance between the points) gives the extensional or compressional strain. Extensometers with baselines as short as a few centimeters are installed in boreholes, and ones with baselines up to tens of meters in tunnels. In both of these, the measurement is made relative to a solid length standard. Much longer instruments use optical methods to measure strain over hundreds of meters. The other class of strainmeter is the volumetric type, in which the change in a volume of fluid is found by sensing the displacement of fluid in and out of a container cemented in a borehole: such an instrument measures the volume strain, or dilatation.

Measuring the rotation vector requires either a stable direction of reference or some way of measuring rotation directly with respect to inertial space. Until recently, no technology existed that do could this with the low noise levels required. A few instruments measure rates of rotation around a vertical axis using Sagnac interferometers, also called laser gyroscopes (Schreiber et al., 2009). These sense the difference in frequency between two light beams propagating in opposite directions around a closed loop.

Seismic data recording

Three specialized requirements for seismological dataloggers are: (1) a high dynamic range, to capture all the signals possible; (2) large amounts of storage, since it may be necessary to record for a long time to capture unpredictable events; and (3) accurate absolute timing, to relate seismic-wave travel times to a common system. In all three areas, the progress of electronics has meant rapidly improving performance at ever-lower cost.

Many seismic dataloggers use specialized systems to provide 24 bits of range (about 5×10^7); this is accomplished by oversampling followed by digital filtering. The amount of storage depends on the sample rate; except in limited situations, a rate of about 200 Hz will capture all seismic data without aliasing. A day of 3-component data with this resolution and sample rate comes to just over 50 Mb, an amount that in the last two decades has gone from requiring specialized and bulky storage to something easily dealt with. Cross-correlation methods can find time delays to a precision of 0.1 of the sample interval, and so the time of each sample point should be known well: for 200 Hz sampling, this requirement would be 500 μs. This too is a level of accuracy that until recently was not easily attained. It can be reached without much difficulty if radio signals from the GPS satellites are available, though providing it over long times in the absence of such a signal still requires expensive equipment. Lower levels of accuracy can be obtained from other radio signals, and, over intervals of a few days, from inexpensive crystal oscillators.

Often, a single system records data telemetered from multiple sensors, forming an array (if the region covered spans only a few wavelengths of the waves being recorded) or network (if larger). The most precise timing then requires corrections for the transmission time of the data (latency) unless a separate datalogger is used for each sensor.

Summary

With over a century of development, inertial seismic sensors are a mature technology, usually capable of recording ground motion much better than it can be modeled. While no single sensor can cover the full range of amplitudes and periods of seismic waves, only a few instruments are needed to provide a faithful record of ground motion. Since other developments in electronics have largely routinized digital recording of seismic data, most users treat seismometers as a "black box" system that can be acquired

and used with little specific expertise: a usually justifiable assumption, though as always it is important to know enough to recognize poor performance when it does occur.

Bibliography

Agnew, D. C., 1986. Strainmeters and tiltmeters. *Reviews of Geophysics*, **24**, 579–624.

Aki, K., and Richards, P. G., 2002. *Quantitative Seismology.* Sausalito: University Science Books.

Berger, J., Davis, P., and Ekström, G., 2004. Ambient Earth noise: a survey of the global seismographic network. *Journal of Geophysical Research*, **109**, B11307, doi:10.1029/2004JB003408.

Bilich, A., Cassidy, J. F., and Larson, K. M., 2008. GPS seismology: application to the 2002 M_W 7.9 Denali Fault earthquake. *Bulletin of the Seismological Society of America*, **98**, 593–606.

Bormann, P. (ed.), 2009. *New Manual of Seismological Observatory Practice.* http://www.iaspei.org/projects/NMSOP.html

Evans, J. R., Hamstra, R. H., Kundig, C., Camina, P., and Rogers, J. A., 2005. TREMOR: a wireless MEMS accelerograph for dense arrays. *Earthquake Spectra*, **21**, 91–124.

Havskov, J., and Alguacil, G., 2004. *Instrumentation in Earthquake Seismology.* Dordrecht: Springer.

LaCoste, L. J. B., 1967. Measurement of gravity at sea and in the air. *Reviews of Geophysics*, **5**, 477–526.

Ringler, A. T., and Hutt, C. R., 2009. Self-noise models of seismic instruments. *EOS Transactions, American Geophysical Union, Fall Meet. Suppl.*, **90**, S23A–1736.

Rodgers, P. W., 1968. The response of the horizontal pendulum seismometer to Rayleigh and Love waves, tilt, and free oscillations of the earth. *Bulletin of the Seismological Society of America*, **58**, 1384–1406.

Rodgers, P. W., 1994. Self-noise spectra for 34 common electromagnetic seismometer/preamplifier pairs. *Bulletin of the Seismological Society of America*, **84**, 222–229.

Schreiber, K. U., Hautmann, J. N., Velikoseltsev, A., Wassermann, J., Igel, H., Otero, J., Vernon, F., and Wells, J. P. R., 2009. Ring laser measurements of ground rotations for seismology. *Bulletin of the Seismological Society of America*, **99**, 1190–1198.

Wielandt, E., 2002. Seismometry (pp. 283–304). In Lee, W. H. K. (ed.), *International Earthquake and Engineering Seismology Part A*. New York: Elsevier Academic Press.

Cross-references

Earthquake, Location Techniques
Earthquakes, Strong-Ground Motion
Free Oscillations of the Earth
Gravimeters
Seismic Noise
Seismogram Interpretation
Seismology, Rotational

SEISMIC MICROZONATION

Fumio Yamazaki, Yoshihisa Maruyama
Graduate School of Engineering, Chiba University, Inage-ku, Chiba, Japan

Definition

Seismic microzonation. The mapping of an area on the basis of various factors that can affect the intensity of ground shaking, such as seismic hazard, geological conditions, and topographical features, so as to account for the effects of local conditions on earthquake-induced damage.

Introduction

Local site conditions affect the intensity of ground shaking, and as a consequence, the extent of earthquake-induced damage. The amplitude, frequency content, and duration of strong ground motion are significantly influenced by local site conditions. A well-known example is the 1985 Mexico City earthquake. Although the fault rupture of the earthquake was about 350 km away from Mexico City, the city sustained catastrophic damage due to the strong amplification of the ground motion by soft soil deposits (Seed et al., 1988). The 1989 Loma Prieta earthquake caused extensive damage in the San Francisco Bay Area. The San Francisco Bay mud significantly influenced the amplitude, frequency content, and duration of ground shaking and resulted in the collapse of the northern portion of the I-880 Cypress Viaduct (Earthquake Engineering Research Institute, 1990; Kramer, 1996). Seismic microzonation provides the basis for site-specific risk analysis, which can assist in the mitigation of earthquake-induced damage.

Methodology

Seismic microzonation typically involves the mapping of predominant periods, soil amplification factors, topographical conditions, liquefaction susceptibility, etc. To draft microzonation maps for a particular region, various data such as existing geological maps, borehole survey data, seismic observation data, and microtremor observation data are collected. Since seismic microzonation entails spatial classification of soil conditions in a small area (e.g., a city), geological data are required for not just a single location, but for many locations. In this regard, geological classification maps are most often used as one of the data sources. However, to classify the target area in a more quantitative manner, actual soil profiles obtained from borehole survey data or seismic observation data are better sources. Unfortunately, in most cases, the borehole survey data and/or seismic observation data available for a small area are insufficient. Thus, microtremor observation data have emerged as a popular source for dense spatial information on site amplification characteristics. Three examples of seismic microzonation are described hereafter.

Example 1. Seismic microzonation based on geomorphological classification maps

Several seismic microzonation studies in Japan have employed geomorphological and geological data from the Digital National Land Information (DNLI), which is a GIS database that covers the whole of Japan with a 1×1 km mesh, to estimate site amplification characteristics (Matsuoka and Midorikawa, 1995; Fukuwa et al., 1998; Yamazaki et al., 2000).

Wakamatsu et al. (2004) drafted the Japan Engineering Geomorphologic Classification Map (JEGM) on the basis of the analysis of local geomorphological features at scales of 1:50,000, and all the attributes were digitized and stored in a GIS database. They recently extended the JEGM to 250 × 250 m grid cells that were categorized into 24 classes on the basis of geomorphological characteristics.

The shear-wave velocity averaged over the upper 30 m (V_s^{30}) is often used as a simplified index of site conditions (Building Seismic Safety Council, 2003). Region-wide site condition maps for California were constructed on the basis of V_s^{30} and the classification of geological units (Wills et al., 2000). The Next Generation of Ground-Motion Attenuation Models (NGA) project was launched in an attempt to collect all publicly available site condition information at strong motion stations. V_s^{30} is used in the absence of site condition information (Chiou et al., 2008). Matsuoka et al. (2006) constructed a nationwide V_s^{30} distribution map using the nationwide shear-wave velocity datasets for Japan, which were obtained from 1,000 K-NET and 500 KiK-net seismic stations and the JEGM.

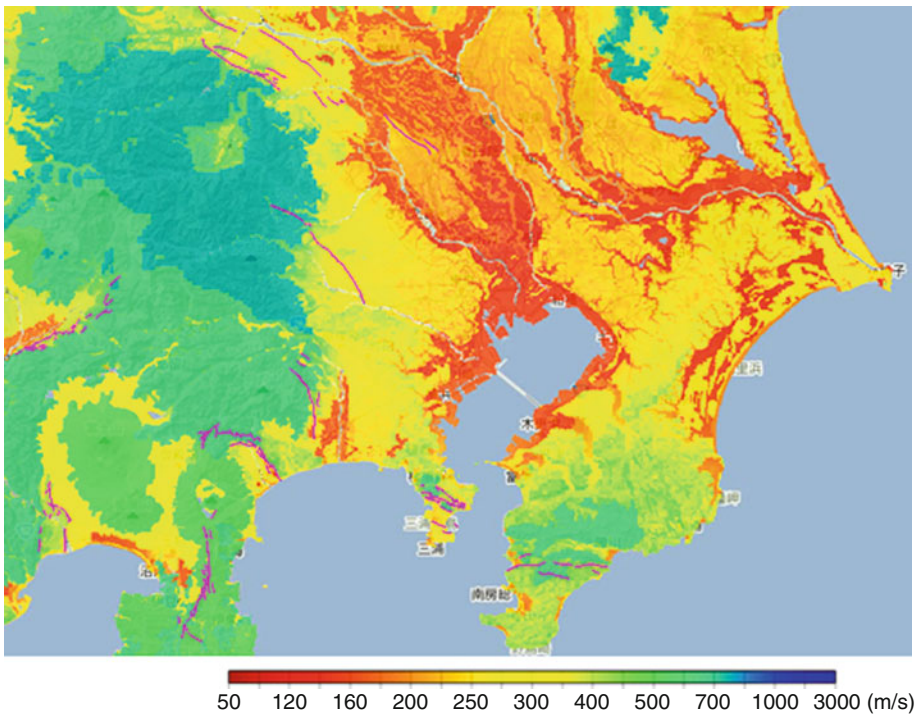
The National Research Institute for Earth Science and Disaster Prevention (NIED), Japan, has developed an open web system that interactively provides seismic hazard maps for Japan; this system is called the Japan Seismic Hazard Information Station (J-SHIS) (Fujiwara et al., 2006). J-SHIS uses the JEGM and V_s^{30} distribution map to draw probabilistic seismic hazard maps for the whole of Japan made by the Headquarters of Earthquake Research Promotion, Japan (Figure 1).

Example 2. Seismic microzonation based on dense borehole data and GIS

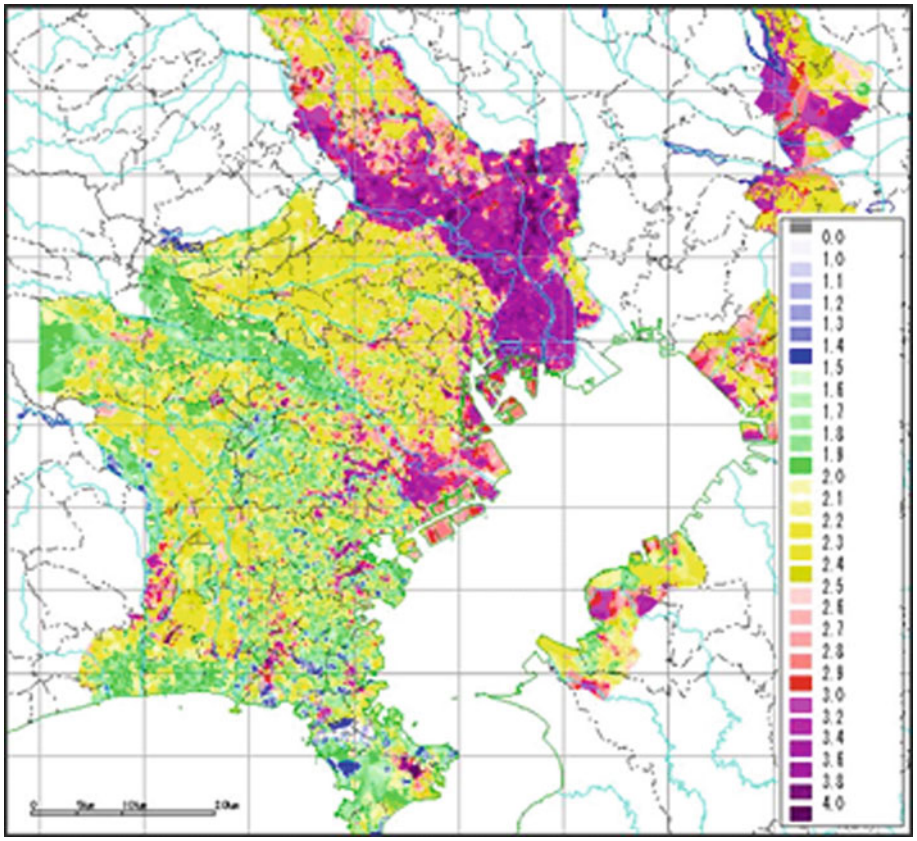
Since 2001, the Tokyo Gas Co., Ltd. has been operating the *Super*-Dense *Real*-time *M*onitoring of *E*arthquakes (SUPREME) system, having about 4,000 seismometers (SI-sensors), in order to control natural-gas supply soon after the occurrence of earthquakes (Shimizu et al., 2006).

This system employs a GIS to interpolate the monitored spectral intensity (SI) values by using subsoil data from 60,000 boreholes. The digitized borehole data specify the location, depths of soil layers, classification of subsurface soil, standard penetration test (SPT) blow counts, surface elevation, and elevation of the ground water table. Thus, microzonation of the area on the basis of individual borehole data is possible. Shear-wave velocities are estimated from an empirical relationship by using the SPT-N values; then, the average shear-wave velocities in the top 20 m of soil at a borehole site are used to estimate the amplification factors of the SI values (Figure 2).

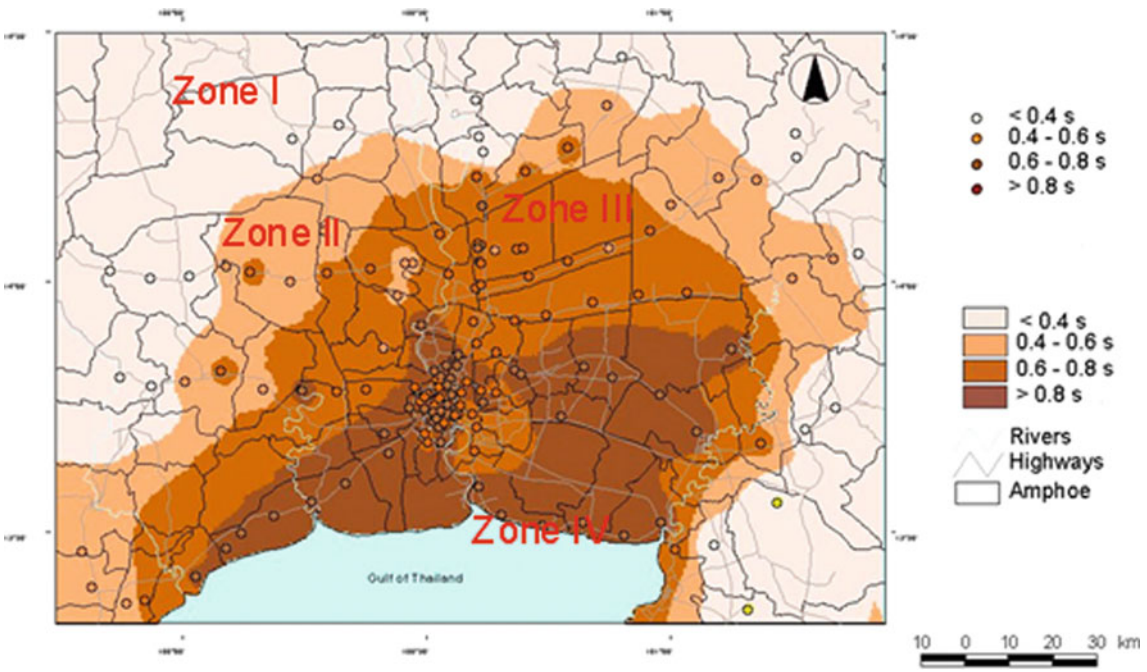
The accuracy of seismic microzonation can be confirmed after several years of operating a dense seismic network by evaluating the seismic records obtained for moderate to small earthquake events occurring in that period.



Seismic Microzonation, Figure 1 V_s^{30} distribution map of Tokyo metropolitan area (http://www.j-shis.bosai.go.jp/).



Seismic Microzonation, Figure 2 Site amplification map of Tokyo and surrounding areas, developed using dense borehole data.



Seismic Microzonation, Figure 3 Microzonation of greater Bangkok area on the basis of variation in predominant period.

Example 3. Seismic microzonation based on microtremor measurements

Microtremor measurements have emerged as a popular tool for determining the dynamic properties of soil layers, and hence, are being widely employed for seismic microzonation. In this method, ambient vibrations (of the order of microns) on the earth's surface are measured. The main sources of these vibrations are traffic and industrial and human activities (Kanai, 1983; Lermo and Chavez-Garcia, 1994). Microtremor measurements can be used to determine the predominant period of vibrations at a site. Nakamura (1989) proposed the horizontal-to-vertical (H/V) spectral ratio method, in which the predominant periods of ground vibrations are determined from the ratio of horizontal and vertical Fourier spectra of the microtremors recorded at a site. Konno and Ohmachi (1998) drafted a map of fundamental periods and amplification factors for the 23 wards of Tokyo on the basis of microtremor measurements carried out at 546 stations.

Tuladhar et al. (2004) drew a seismic microzonation map for the greater Bangkok area, Thailand, on the basis of microtremor observations carried out at 150 sites. The predominant periods of these sites were obtained by using the H/V method. The estimated predominant periods were validated by comparing them with the transfer functions obtained from one-dimensional wave-propagation analysis conducted at eight sites. According to the variation in the predominant period of the ground, the greater Bangkok area was classified into four zones as follows: Zone I (period less than 0.4 s), Zone II (0.4–0.6 s), Zone III (0.6–0.8 s), Zone IV (longer than 0.8 s). Figure 3 illustrates the microzonation of the greater Bangkok area on the basis of variation in the predominant period.

Summary

The objectives and methodologies to perform seismic microzonation are described and some examples are presented. The three major methods introduced to achieve seismic microzonation are the uses of geomorphological classification maps, dense borehole datasets, and microtremor measurements. The results of seismic microzonation are compiled for a GIS to draft microzonation maps and they can be used to predict ground motions during disastrous earthquakes and thus can assist in the mitigation of earthquake-induced damage.

Bibliography

Building Seismic Safety Council (BSSC), 2003. *The 2003 HEHRP Recommended Provisions for the Development of Seismic Regulations for New Buildings and Other Structures*. Washington, DC: FEMA.

Chiou, B., Darragh, R., Gregor, N., and Silve, W., 2008. NGA project strong-motion database. *Earthquake Spectra*, **24**(1), 23–44.

Earthquake Engineering Research Institute, 1990. Loma Prieta earthquake reconnaissance report. *Earthquake Spectra*, **6**.

Fujiwara, H., Kawai, S., Aoi, S., Morikawa, N., Senna, S., Kobayashi, K., Ishii, T., Okumura, T., and Hayakawa, Y., 2006. National seismic hazard maps of Japan. *Bulletin of the Earthquake Research Institute, University of Tokyo*, **81**, 221–231.

Fukuwa, N., Arakawa, M., and Nishizaka, R., 1998. Estimation of site amplification factor using Digital National Land Information. *Journal of the Structural Engineering*, **44**(B), 77–84 (in Japanese).

Kanai, K., 1983. *Engineering Seismology*. Tokyo: University of Tokyo Press, pp. 128–139.

Konno, K., and Ohmachi, T., 1998. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bulletin of the Seismological Society of America*, **88**(1), 228–241.

Kramer, S. L., 1996. *Geotechnical Earthquake Engineering*. Englewood Cliffs: Prentice Hall.

Lermo, J., and Chavez-Garcia, F. J., 1994. Are microtremors useful in site response evaluation? *Bulletin of the Seismological Society of America*, **84**(5), 1350–1364.

Matsuoka, M., and Midorikawa, S., 1995. GIS-based integrated seismic hazard mapping for a large metropolitan area. In *Proceedings of the Fifth International Conference on Seismic Zonation*, Nice, France, Vol. II, pp. 1334–1341.

Matsuoka, M., Wakamatsu, K., Fujimoto, K., and Midorikawa, S., 2006. Average shear-wave velocity mapping using Japan Engineering Geomorphologic Classification Map. *Journal of Structural Engineering and Earthquake Engineering (Japan Society of Civil Engineers)*, **23**(1), 57s–68s.

Nakamura, Y., 1989. A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly Reports of the Railway Technical Research Institute*, **30**, 25–33.

Seed, H. B., Romo, M. P., Sun, J. I., Jaime, A., and Lysmer, J., 1988. The Mexico earthquake of September 19, 1985 – relationships between soil conditions and earthquake ground motions. *Earthquake Spectra*, **4**, 687–729.

Shimizu, Y., Yamazaki, F., Yasuda, S., Towhata, I., Suzuki, T., Isoyama, R., Ishida, E., Suetomi, I., Koganemaru, K., and Nakayama, W., 2006. Development of real-time control system for urban gas supply network. *Journal of Geotechnical and Geoenvironmental Engineering, ASCE*, **132**(2), 237–249.

Tuladhar, R., Yamazaki, F., Warnitchai, P., and Saita, J., 2004. Seismic microzonation of the greater Bangkok area using microtremor observations. *Earthquake Engineering and Structural Dynamics*, **33**(2), 211–225.

Wakamatsu, K., Matsuoka, M., Hasegawa, K., Kubo, S., and Sugiura, M., 2004. GIS-based engineering geomorphologic map for nationwide hazard assessment. In *Proceedings of the 11th International Conference on Soil Dynamics & Earthquake Engineering and 3rd International Conference on Earthquake Geotechnical Engineering*, Vol. 1, pp. 879–886.

Wills, C. J., Petersen, M., Bryant, W. A., Reichle, M., Saucedo, G. J., Tan, S., Taylor, G., and Treiman, J., 2000. A site-conditions map for California based on geology and shear-wave velocity. *Bulletin of the Seismological Society of America*, **90** (6B), S187–S208.

Yamazaki, F., Wakamatsu, K., Onishi, J., and Shabestari, K. T., 2000. Relationship between geomorphological land classification and site amplification ratio based on JMA Strong Motion Records. *Soil Dynamics and Earthquake Engineering*, **19**(1), 41–53.

Cross-references

Earthquakes, Intensity
Earthquakes, Strong-Ground Motion
Seismic Hazard
Seismic Zonation
Seismicity, Intraplate
Seismology, Global Earthquake Model

SEISMIC MONITORING OF NUCLEAR EXPLOSIONS

Paul G. Richards[1], Wu Zhongliang[2]
[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA
[2]Institute of Geophysics, China Earthquake Administration, Beijing, China

Introduction

The original development of nuclear weapons, and their first use in 1945, was followed by several decades of further weapons development in which more than 2,000 nuclear test explosions were conducted. About 500 of these were carried out in the atmosphere, mostly in the 1950s and 1960s. They generated radioactive fallout that was detected worldwide with some regional concentrations, and aroused widespread public opposition to nuclear testing. A few nuclear tests were carried out underwater and in space. The great majority, about 1,500, were conducted underground in ways that greatly reduced fallout – the first of them in 1957, in Nevada, USA – generating signals that have been intensively studied by seismologists. Hundreds of these individual nuclear tests consisted of multiple nuclear devices and exploded almost simultaneously.

A ban on nuclear testing in the atmosphere, underwater, or in space, was negotiated and went into effect in 1963 between the USA, the USSR, and the UK. Known as the Limited Test Ban Treaty (LTBT), it has since been ratified or acceded to by more than a hundred countries. Though France and China did not sign, and China carried on with nuclear testing in the atmosphere up to 1980, eventually both these countries came to abide by its terms.

The concept of a Comprehensive Test Ban Treaty (CTBT) emerged in the 1950s, intended as a restraint upon nuclear weapons development. It was debated in many forums for more than 40 years, and finalized in terms of specific treaty text in September 1996. But this treaty is not in effect (as of 2010), due to continuing debate in specific countries that have not ratified this treaty, and whose ratification is needed as a condition for the CTBT to enter into force. They include India, North Korea, and Pakistan (not signed or ratified); and China, Israel, and the United States (signed but not ratified). Those countries that have signed the treaty are effectively adhering to a moratorium on nuclear testing. They include the five countries recognized as nuclear weapons states by the Non-Proliferation Treaty of 1968. Listing them in the order in which they acquired nuclear weapons capability, these are the USA, the USSR (whose CTBT obligations have been assumed by Russia), the UK, France, and China. The two countries that by far have conducted the most nuclear test explosions – the USA with 51% of the world total, and the USSR/Russia with 35% – ended nuclear testing in the early 1990s. See Yang et al. (2003) for lists of nuclear explosions conducted in the twentieth century, and Bennett et al. (2010) for a relevant database and seismic waveforms. Since 1996, the only nuclear explosions (as of 2010) have been those conducted by India and Pakistan (in May 1998), and by North Korea (in October 2006, and May 2009).

Seismic monitoring of nuclear explosions has been an important activity ever since the first nuclear test in July 1945 in New Mexico. Such monitoring is driven by two different objectives that have engaged a range of different institutions and organizations. The first objective, which dominated for the early decades of nuclear testing up to the early 1990s when nuclear explosions were being conducted on average about once a week, was to acquire basic information about military weapons being tested, especially if (from the point of view of the monitoring organization) the tests were being carried out by a potential adversary. Relevant questions were: what countries had nuclear weapons programs, developed to the level of carrying out nuclear explosive tests? And how big were these explosions? The second objective, which has become important in recent decades, has been in the context of a major initiative in nuclear arms control, namely, to achieve confidence in the capability to monitor compliance with a CTBT, recognizing that many countries considering whether or not to support such a treaty and to be bound by its terms, would need to have confidence in the monitoring system to some adequate degree. Given that monitoring cannot be done all the way down to zero yield, evaluation of progress toward this second objective entails questions such as: down to what small size can nuclear explosions be detected, and identified, and attributed with high confidence? And what are the specific capabilities of different types of monitoring program, applied to different parts of the world, to catch evidence of a nuclear test, should one occur?

Seismology is the most effective technology for monitoring nuclear tests carried out underground, which is the one environment that was not covered by the LTBT, and which is also the hardest of the environments to monitor. The importance of achieving the two objectives stated above has shaped modern seismology itself, in that much of the funding that has led to the facilities and bodies of knowledge now used widely in seismological research (including studies of seismic hazard), were stimulated by government programs intended to improve capabilities for seismic monitoring of nuclear explosions. These facilities and methods include high-quality Seismic Instrumentation, global networks that monitor for earthquakes as well as explosions, quantitative methods of characterizing seismic sources (various magnitude scales, the moment tensor), theoretical understanding of seismic wave propagation in Earth models of increasing and more realistic complexity, our knowledge of the Earth's internal structure, and methods of seismic signal detection and interpretation.

The technical capability to monitor explosions, or a perceived lack of such capability, has played a role in the development of policy options on weapons testing

and/or arms control and the content of international treaties. A key technical question arising in debates has been: down to what value of yield can monitoring be accomplished – and with what level of confidence? Seismologists claim now that there is no fundamental technical problem with monitoring explosions down to 1 kt, even if determined efforts at evasion must be considered. But there have been assertions that it is possible to muffle and thus hide (or confuse the procedures for identifying) the seismic signal, even from a substantial underground explosion at the level of ten kilotons or more. These latter assertions do not appear plausible after review of the technical difficulties; but, as assertions, one finds that they continue to survive.

Seismic monitoring for underground nuclear explosions must be done with recognition of the great variety and number of earthquakes, chemical explosions, and other nonnuclear phenomena that generate seismic signals every day. Efforts to sort out and identify signals from underground nuclear explosions in the midst of signals from these other phenomena have made great progress since they commenced in the 1950s, and improvements in monitoring capability will surely continue to be made.

Sections below describe basic properties of earthquake and explosion signals, and different steps in seismic monitoring for nuclear explosions. A review is given of methods used for decades in the era when thousands of kilometers separated nuclear weapons testing activity and monitoring stations, when nuclear weapons testing was commonplace and there was little incentive to hide testing activity. Descriptions are then given of modern methods that monitor for very small explosions and the possibility of tests conducted in ways intended to evade discovery. A description is given of so-called "problem events" that were important is developing effective and in some cases new discriminants; and finally a brief summary is given of monitoring capabilities, as of 2010, emphasizing the utility of data and data products from the International Monitoring System and its associated International Data Centre that are operated today by the CTBT Organization, headquartered in Vienna, Austria.

Basic properties of earthquake and explosion signals

Seismic monitoring for underground nuclear explosions has to face the reality of hundreds of earthquakes, chemical explosions, and other nonnuclear phenomena, generating seismic signals daily that will be recorded at multiple stations by any effective monitoring network. But after decades of effort, an extensive infrastructure of national and international agencies now sorts out and identifies the signals from earthquakes, chemical explosions, and the occasional underground nuclear explosion. Modern methods of nuclear explosion monitoring are vastly more capable than they were when this work began in the late 1950s. The improvements have mostly been steady as data quality and quantity from monitoring networks increased, but with occasional jumps in capability as new types of analysis were validated.

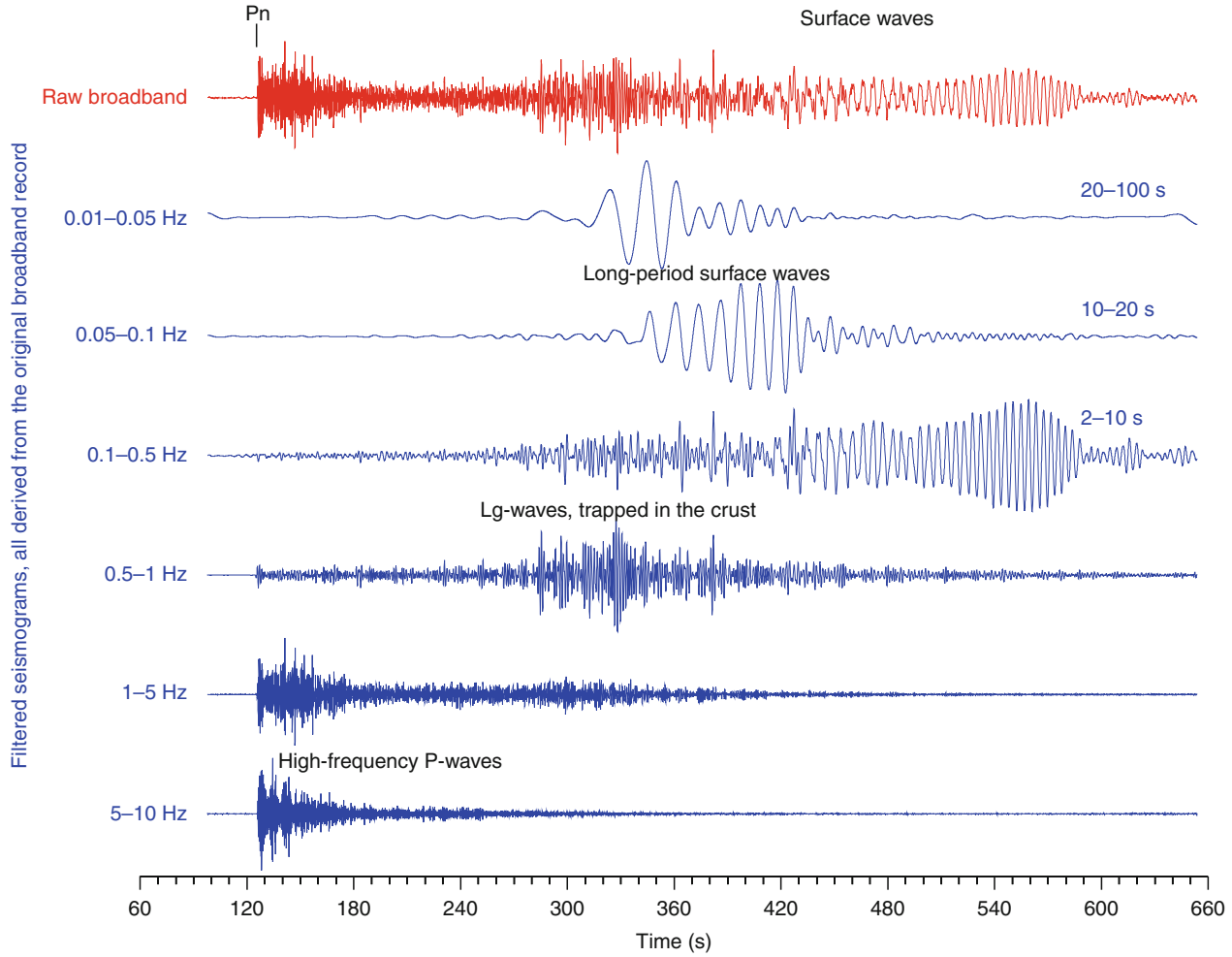
Seismic signals are traditionally grouped into teleseismic waves and regional waves, depending on the distance at which they are observed. Teleseismic waves propagate either as Body Waves through the Earth's deep interior, emerging with periods typically in the range 0.3–5 s at distances greater than about 1,500 km, or as Surface Waves, analogous to the ripples on the surface of a pond, with periods of about 15–100 s.

Teleseismic waves were the basis of most US monitoring of foreign nuclear tests prior to 1987. Teleseismic body waves are further subdivided into *P*-waves and *S*-waves. *P*-waves, which are the fastest-traveling seismic waves and are therefore the first to arrive, are excited efficiently by explosions: earthquakes tend to excite *S*-waves and surface waves more efficiently.

For subkiloton explosions, teleseismic signals can be too weak for detection at distant stations and monitoring then requires regional signals. Regional waves are of several types, including *P*-waves and *S*-waves, all propagating only at shallow depths (less than 100 km below the Earth's surface) with periods as short as 0.05 s (frequencies as high as 20 Hz, i.e., cycles per second). Regional waves reach distances up to 1,000 km and sometimes beyond, depending on source size and whether the propagation path is an attenuating one, or not. They are regional also in the sense that they have speeds and attenuation properties that vary according to details of local structures in the Earth's crust and uppermost mantle, so they can vary from place to place within continents and oceans.

Figure 1 shows a regional seismogram of a Soviet underground nuclear explosion in Kazakhstan recorded in July 1989 at a distance of slightly less than 1,000 km by a high-quality station in northwestern China. The original recording is shown in red. Different signals derived from it are shown in blue, each of them filtered to pass information in a particular band of frequencies.

Seismologists characterize the size of seismic signals by means of logarithmic magnitude scales (see *Earthquake, Magnitude*), with each scale based on a different type of seismic wave. A magnitude scale using teleseismic surface waves was first described in the 1930s based on the logarithm (to the base 10) of amplitude of maximum ground displacement due to surface waves with periods about 20 s. It is known as the M_s scale. Another widely used magnitude scale is that based on the amplitude of teleseismic *P*-waves. Known as m_b, it entails measurement of ground motion at about 1 s period. As part of the assignation of M_s and m_b values, for a particular seismic event as recorded at a particular station, a standard correction is applied to account for the distance between the source and the receiver at which the data was obtained. Magnitudes range from about -3 for the smallest observable microearthquakes, up to above 8 for the largest earthquake. A 1 kt underground explosion has an m_b roughly about 4, and each year there are about 7,500 shallow earthquakes worldwide with $m_b \geq 4$ (Ringdal, 1985). Although use of



Seismic Monitoring of Nuclear Explosions, Figure 1 The seismogram recorded at station WMQ in northwestern China, for an underground nuclear explosion on July 8, 1989 in Kazakhstan at a distance of almost 1,000 km, is shown in red (*top*). Filtered versions of the original trace in different frequency bands are shown in *blue*. Time in seconds at *bottom* is with respect to the time the explosion occurred. Different types of seismic wave propagate at different frequencies, and hence their ground motions show up in different bands. *P*-waves, in this case the regional wave called *Pn* that travels in the uppermost mantle, arrive about 120 s after the explosion at this distance, involving short-period (high frequency) motions. Long-period surface waves can be seen in the top two *blue* traces. Some surface waves arrive up to 600 s after the explosion at this distance and, thus, travel as much as five times slower than *P*-waves. *S*-waves (weak in this example) are shear waves, traveling slower than *P* waves. A high-frequency wave marked as *Lg*, which is often the largest wave at regional distances from an earthquake but is only weakly excited by explosions, is dominated by shearing motions and is largely trapped in the Earth's crust. The amplitude of ground motion in the longest period band is less than 2% the amplitude in the short period band from 1 to 5 Hz. (Adapted from work of W.-Y. Kim.)

seismic moment has superseded use of m_b and M_s in much of modern seismology and magnitude is only an empirical estimator of seismic event size, magnitude scales are still often used in discussion of seismic monitoring because this a practical way to relate that discussion directly to properties of signal strength. For example, monitoring capability is often characterized in terms of contour maps or shaded maps indicating the magnitude levels down to which detection or identification is deemed possible with given resources, such as a particular network. We conclude this article with such a map (see Figure 8). Explosion energy is measured in kilotons. A kiloton is formally defined as a trillion calories, and is roughly the energy released by exploding a thousand tons of TNT.

The different steps in explosion monitoring

Nuclear explosion monitoring entails a series of steps, beginning with *detection* of signals (did a particular station detect anything?) and *association* (can we gather all the different signals, recorded by different stations, that originate from the same "event"?). The next steps involve making a *location* estimate and an *identification* (did it have the characteristics of an earthquake, a mining blast,

a nuclear weapon test?). Then follow the steps of *yield estimation* (how big was it?) and *attribution* (if it was a nuclear test, what country carried it out?).

Detection

Concerning detection, nuclear explosion monitoring is often done with arrays of sensors, deployed as a group spread out over an area about 10 km across (or less), that facilitate methods to enhance signal-to-noise ratios. This is done typically by stacking signals from independent sensors, often with appropriate delays to increase signal strength and reduce noise. Array data can also give estimates of the direction from which signals are arriving.

In the *evaluation* of detection capability, one of the key concepts widely used in seismology is the *magnitude of completeness*, which means that *all* events above this magnitude can be recorded by the monitoring system. Transferring from magnitude to yield, one infers the capability for detecting nuclear tests (NAS, 2002). Practically, however, one of the often-cited expressions of monitoring capability is the *magnitude threshold*, above which 90% of the seismic events can be detected at more than three stations, the least number of stations for routine location.

Association

Association is the effort to identify those sets of signals, from different stations, which all originate from the same seismic event. It is one of the hardest steps in practice, particularly when multiple seismic sources around the world are active at the same time, resulting in signals from different events that are interlaced in the waveforms recorded by each station. In such cases, array data can be helpful in resolving which signals correspond to which event.

Location

To obtain a location estimate, typically the arrival times of various seismic waves are measured from the recorded waveforms such as shown in Figure 1. They are used to find four parameters: latitude, longitude, depth, and origin time. In this work, it is necessary to know the travel time from any hypothesized source location to any particular seismographic station for any type of seismic wave that the station might observe. In practice, locating seismic events accurately on a global basis (say, to within 10 km of their true location) using sparse networks (stations several hundred kilometers apart) requires extensive efforts in station calibration. Thus, it is important to include path-specific travel-time corrections to standard travel-time models to account for lateral variations of Earth structure (Murphy et al., 2005; Myers et al., 2010). Many authors have shown that greatly improved precision of location estimates can be achieved for a given region if seismic events are located in large numbers – preferably thousands of them or more, all at the same time – rather than one at a time (Richards et al., 2006; Waldhauser and Schaff, 2008).

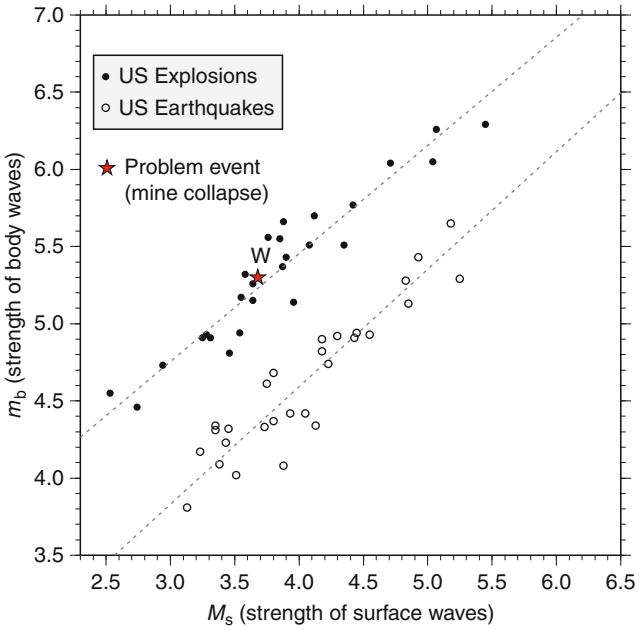
Methods of identification

Identification of the nature of a seismic source on the basis of its seismic signals – that is, making a determination from seismograms as to whether it could be a nuclear explosion, or a natural earthquake, or a mine blast, or something more exotic such as a bolide impacting our planet and exploding in the atmosphere – is a large subject in view of the many possibilities. See for example, Richards (1988), OTA (1988), Dahlman et al. (2009), and Bowers and Selby (2009). Seismic events generate many different types of seismic wave, in various different frequency bands as shown in Figure 1, and different types of seismic source generate a different mix of seismic waves. We can make an analogy here with sound waves, and the capability of the human ear and brain to analyze them. A deep bass voice, a gunshot, a whistle, and rolling thunder, constitute a set of sound sources that are easily distinguished from each other on the basis of their different frequencies, their emergent or impulsive nature, and their duration. It is the mix of information in both the time domain and the frequency domain that is effective.

Seismic methods for discriminating between earthquakes and explosions are based on interpretation of the event location (including its depth); on the relative excitation of a variety of body waves and surface waves; and on properties of the signal spectrum associated with each of these two different types of source. Within these three broad categories, many different methods have been tried, with various degrees of success. As the capabilities of each method are probed, the question of interest is often: "Down to what size of seismic event, does this method of discrimination work?" In some cases, discrimination is unambiguous even at very small event size. (For example, however small an event, it may be presumed to be an earthquake if it is located at a depth greater than 15 km below the Earth's surface. Even a small event will attract attention if it occurs in an area that is geologically stable that for decades has had no seismic activity.)

The most useful methods for discrimination can be listed as follows:

- Interpretation of the location: Is the event in a seismic or an aseismic area? Below the floor of an ocean? At depth below a continent? There is an important role here for common sense: seismic events in Canada tend to attract less attention from western monitoring agencies than such events in North Korea (though a seismic event in the middle of the Canadian Shield would still attract attention and intensive study).
- Relative amplitude of body waves and surface waves. This can be studied by plotting the event of interest on an M_s: m_b diagram, as shown in Figure 2. The surface-wave amplitude is read typically from signals with period about 20 s, and the body-wave amplitude at about 1 s period. (Though effective for large enough



Seismic Monitoring of Nuclear Explosions, Figure 2 An M_s: m_b diagram from Bowers and Walter (2002). It can be seen here that for seismic events of the same M_s value, earthquakes have a significantly smaller m_b value than do the explosions. The offset is about 0.8 m_b units, at $M_s = 5$. Because magnitudes are based on logarithmic scales, and $10^{0.8} \sim 6$, it follows that at frequencies near those at which body wave magnitude is measured (about 1 Hz), the *P*-waves from an underground nuclear explosion are about 6 times larger than such waves from an earthquake having the same strength of surface waves. Also, indicated by the red star are the body-wave and surface-wave magnitudes of an interesting but fortunately rare event, a large mine collapse with *P*-wave magnitude greater than 5. This event, which plots with the explosion population, is discussed further below – see Figures 6 and 7.

events, an explosion with m_b much below 4.5 may not have large enough surface wave signals at teleseismic distances to apply this method dependably.)

- Use of the observed "first motion" of the ground. Is the initial *P*-wave motion of the ground indicative of compression radiated to all directions from the source, leading to upward motions, as would be the case for a simple explosion? Or, are dilatations recorded at some azimuths, leading to downward motions, as would sometimes be expected from earthquakes but not from explosions?

The methods described so far in this section have concerned the use of teleseismic signals, which can be used to monitor effectively for high magnitudes, and on down to somewhere in the magnitude range from 4.0 to 4.5. Since the early 1990s, there has been growing recognition of the merits of regional waves, to monitor down to far lower magnitudes, often well below magnitude 3. The method is based upon the general observation that explosion signals, when compared to earthquakes, have much stronger *P*-waves at high frequency, whereas those from earthquakes have stronger *S*-waves (and surface waves).

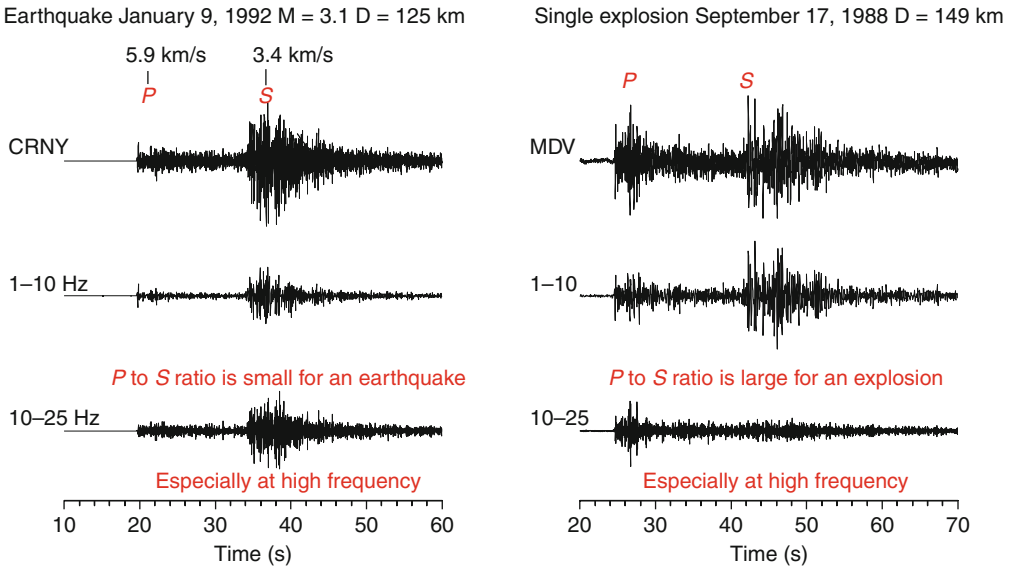
This modern method is being studied with frequencies in the range 0.5–20 Hz. and sometimes even higher. An example is shown in Figure 3 comparing regional signals of a very small earthquake and a small explosion. The method has been demonstrated even down to around m_b 2.

As an important example of this development, Figure 4 shows the results of an analysis of the *P*-wave and *S*-wave spectra, pertinent to identifying the very small underground nuclear explosion conducted by North Korea on October 9, 2006, and the larger test nearly 3 years later on May 25, 2009. The smaller explosion took place at 0135 h (GMT) and by 0706 h the US Geological Survey (USGS) had issued a report based on seismic signals from 20 stations around the world including sites in China, South Korea, Russia, Japan, Kazakhstan, Kyrgyzstan, Alaska, and Nevada. Its magnitude, about 4, indicated a sub-kiloton yield (see Koper et al., 2008, who discuss the uncertainty of estimating yield in view of the variability of seismic signal excitation for shots of different depth). But from such teleseismic signals, the nature of the event was difficult to distinguish from an earthquake. Fortunately, discrimination for events such as this is often very clear, provided high-quality regional data is available.

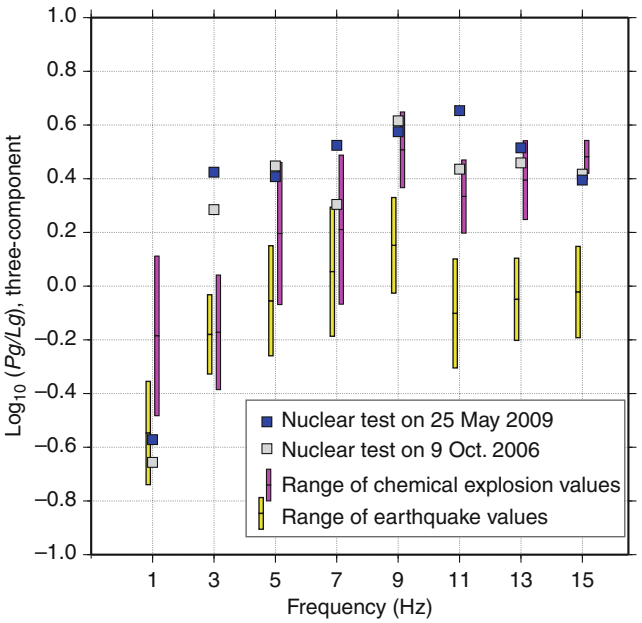
In this analysis, the original seismograms from station MDJ, located in China, are filtered in eight narrow frequency bands as illustrated in blue in Figure 1, but this time with bands centered on each of the frequencies from 1, 3, 5, 7, 9, 11, 13, to 15 Hz as indicated for the horizontal axis in Figure 4. The amplitudes of the *Pg* and *Lg* waves are measured in each narrow band, the amplitude ratio is formed (the "spectral ratio"), and the quantitative comparison can begin. Figure 4 shows how this ratio varies with frequency for the set of eight earthquakes, and for the set of four small chemical explosions. The ratio differs for these two populations as frequency rises, and the separation between them is very clear at high frequencies (from 9 to 15 Hz in this case). It is also clear that the spectral ratios of the signal recorded for the events of 2006 and 2009 are like those of the known chemical explosions.

This successful seismic discriminant based upon regional waves is important in enabling monitoring capability to be extended down to lower magnitudes. In practice, there is often very little difference between the magnitude thresholds for detection (at enough stations to enable a useful location estimate), and identification, since so many regions of the Earth are now monitored to low magnitude for earthquakes as part of investigations into seismic hazard. It may take only one regional seismogram to enable discrimination to be carried out with high confidence (provided the recording is of adequate quality, and is for a station that has an archive of signals from previous known earthquakes and explosions).

Along with the use of regional seismic waves and their spectral ratios at 5 Hz and higher, another discriminant turning out to be successful at distinguishing between



Seismic Monitoring of Nuclear Explosions, Figure 3 Typical vertical-component records from an earthquake and an explosion. Traces plotted are: unfiltered (*top*), low-frequency bandpass filtered (*middle*), and high-frequency bandpass filtered (*bottom*). (From Kim et al., 1993.)



Seismic Monitoring of Nuclear Explosions, Figure 4 Spectral ratios are shown, for the two nuclear explosions carried out by North Korea in 2006 and 2009, as measured from waveforms recorded at station MDJ in China (distance, about 370 km). They are compared with these ratios for a small group of earthquakes, and another group of chemical explosions, all in the vicinity of North Korea's nuclear test site. Colored bars represent ±1 standard deviation in the ratios for chemical explosions (*magenta*), and small earthquakes (*yellow*). The spectral ratios for events in North Korea on October 9, 2006, and on May 25, 2009, are both explosion-like. (Courtesy of Won-Young Kim.)

earthquakes and explosions is the use of observed seismic waveforms to make estimates of the set of forces that appear to be acting at the seismic source. The set of forces here is quantified by what seismologists call the *moment tensor*. As shown by Ford et al. (2009) from study of numerous earthquakes and underground explosions, seismic events separate into specific populations as determined by the way their moment tensors behave – whether they are more representative of the all-around (isotropic) features of an explosion, or of the type of shearing motions more typical of an earthquake.

In general for underground tests, seismic data alone cannot distinguish between nuclear explosions, and chemical explosions in which all the material making up the explosive is fired within less than about a tenth of a second. But such chemical explosions, if large, are very rare. In the case of the two North Korea tests, both of which were announced as nuclear, objective evidence for the nuclear nature of the 2006 explosion came from several different detections of radionuclides that are diagnostic of a nuclear explosion. Such radionuclides were not detected from the 2009 explosion, which, however, was so large as to be implausible as a chemical explosion, since it would have to have consisted of literally thousands of tons of explosives.

Yield estimation

Yield estimation was of particular importance in the years following 1974 when a bilateral treaty between the USA and the USSR was negotiated, intended to go into effect in 1976. This was the Threshold Test Ban Treaty (TTBT), limiting the size of underground nuclear explosions

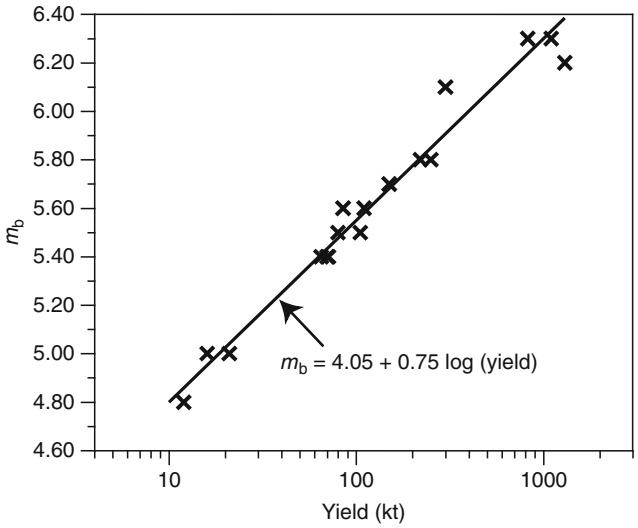
conducted by these two countries to a yield of not more that 150 kt. The TTBT proved contentious, with each side sending the other several inquiries asserting that the agreed-upon limits had possibly been exceeded (Timerbaev, undated). But this treaty was finally ratified in 1990, and has become less important since the CTBT was finalized and a nuclear testing moratorium by the signatory countries began in 1996. Yield estimation is however still important as an exercise in the interpretation of signals from the few underground explosions since that date, specifically those of India and Pakistan in 1998, and of North Korea in 2006 and 2009.

For a few tens of underground nuclear explosions, most of them at the Nevada Test Site, the yield has been announced by the agency conducting the test. It has therefore been possible to calibrate observed seismic magnitudes for these tests against the announced yields, and an example is given in Figure 3 using m_b values and yields reported for Nevada explosions in tuff and rhyolite.

The line $m_b = 4.05 + 0.75$ log (Yield) fits the data well (yield in kilotons). Such a calibration curve can be applied to obtain a seismic yield estimate for Nevada explosions with unannounced yield. But it requires correction, prior to its use in obtaining a seismic yield estimate for an explosion at a different site. This must be done, to allow for physical and geological differences between the sites. For example, in different rock types there can be different efficiencies in the coupling of nuclear yield into seismic energy; and differences in the propagation efficiencies as seismic waves travel out from the source of interest, as compared to seismic signals from a Nevada explosion. In this connection, it is of interest to note m_b and yield for the US nuclear explosion LONGSHOT (conducted in 1965 in the volcanic breccias of an Aleutian island). The m_b value is 5.9, corresponding to a yield of about 300 kt. if the Nevada curve of Figure 5 is applied directly. But the announced yield for LONGSHOT is 80 kt. One way to obtain a calibration curve for the Aleutians is therefore to add a correction of about 0.4 m_b units to the Nevada values of m_b at a given yield, before the curve of Figure 5 is used to supply a seismic yield estimate in this new location. This m_b correction, for a site differing from that where a calibration curve is directly available, is called the *bias*. If the bias correction is not applied, then a Nevada magnitude–yield curve can give too high a seismic yield estimate for a non-Nevada explosion.

Note that the Nevada Test Site is in a region of active tectonics, with significant episodes of volcanism in the last few million years, resulting in high temperatures within the upper mantle, and thus anomalous attenuation of seismic waves propagating through the hot and partially molten upper layers of the Earth, 100 or 200 km in thickness beneath the Nevada Test Site. Such propagation through an attenuating medium is presumed to be a contributing cause of bias.

The existence of m_b bias has long been known in seismology in connection with what is called "station bias." By this term is meant the systematic difference between



Seismic Monitoring of Nuclear Explosions, Figure 5 Seismic magnitude m_b vs. announced yield, for 17 Nevada Test Site nuclear explosions in tuff and rhyolite. The straight line here, which fits the data quite well, can be used to make a yield estimate of other events at this test site, in similar rock, if the seismic magnitude is known. (Data from Nuttli, 1986.)

mean m_b values (obtained for a particular seismic event by averaging reported m_b from seismometers all over the globe), and m_b reported by just one station. For example, the station BMO in Oregon (another region of active tectonism) has reported m_b values that for a given earthquake are typically about 0.3 units below the global average; and station KJN in Finland (in a stable shield region) reports values about 0.15 m_b units higher than the average. Their station bias values are thus −0.3 and +0.15, respectively. Station bias values commonly range over ±0.4 m_b units, so it may be expected that source region bias (which is what must be applied when a standard m_b – yield curve is used for different source regions) will also range over about 0.8 m_b units.

The nuclear weapons test site of the USSR that conducted the most underground nuclear explosions was near the city of Semipalatinsk, in northeastern Kazakhstan. Several multi-megaton underground explosions were conducted on Russia's Novaya Zemlya island test site, far to the north of continental Eurasia (see Khalturin et al., 2005). But these were all prior to the intended date of entry-into-force of the TTBT (March 1976). After that date, the magnitude of the largest underground tests at Semipalatinsk rose higher and higher over several years, with some magnitudes exceeding 6.1. Such magnitudes, according to the Nevada Test Site formula discussed above, $m_b = 4.05 + 0.75$ log (Yield), implied yields great than 500 kt, far in excess of the TTBT limit (150 kt). Intensive discussion in political and technical arenas ensued with stronger and stronger evidence accumulating to indicate a substantial test site bias between the Nevada and Semipalatinsk test Sites. For example, it

was of great interest that teleseismic signals from the largest underground explosions from these two tests, if recorded at the same station in a shield region, looked significantly different. The teleseismic *P*-wave from a large underground explosion at the site in Kazakhstan would routinely have frequency content at the 5 Hz level and sometimes higher (Der et al., 1985). The signal from Nevada would not contain such high frequencies. It was as if the signal from Nevada had passed through some type of filter, which of course would reduce its amplitude. Correcting for that effect would mean that the appropriate relation between magnitude and yield for an underground nuclear explosion at Semipalatinsk had the form

$$m_b = 4.05 + \text{bias} + 0.75 \log (\text{Yield}),$$

and Ringdal et al. (1992) and Murphy (1996) among many others concluded that the appropriate formula relating teleseismic *P*-wave magnitude and yield at Semipalatinsk should be this equation with a bias of 0.4. Support for this conclusion came from many arguments (see Richards, 1988 for a review). But in the political realm, the most persuasive was the very practical one associated with a Joint Verification Experiment of September 14, 1988, in which a team from the USA at the Semipalatinsk Test Site was allowed to make close-in measurements (within a few tens of meters) of a large Soviet underground nuclear explosion, in particular of the speed and extent of the shock wave it sent out into rock near the source at that test site. From such shock measurements, a reliable non-seismic method provided an accurate yield estimate (it was in the range 100–150 kt). Stations around the world provided measurements teleseismically, giving a seismic magnitude around 6.1 – comparable with the largest magnitudes of Semipalatinsk explosions since 1976, indicating that they too had been conducted in a way that respected the 150 kt limit of the TTBT. A reciprocal Joint Verification Experiment had been conducted at the Nevada Test Site, on August 17, 1988 with a Russian team making its own close-in measurements of the shock wave from a large US underground nuclear test intended to be in the range 100–150 kt. According to many news reports, the yield of this explosion slightly exceeded 150 kt. Timerbaev (undated) and news reports give it as 180 kt.

Problem events

The work of monitoring – for both earthquakes and explosions – is done in practice by hundreds of professionals who process the vast majority of seismic events routinely, and who also look out for the occasional events that, in the context of monitoring for the possibility of underground nuclear explosions, exhibit interesting characteristics, and which may then become the subject of special study.

These special events have stimulated the development of effective new discrimination techniques and a better appreciation of overall monitoring capability. Examples include a mine collapse in 1989 in Germany and two such collapses in 1995, in the Urals (Russia) and in Wyoming (USA); a small earthquake of magnitude 3.5 and its smaller aftershock in 1997 beneath the Kara Sea near Russia's former nuclear test site on Novaya Zemlya; and two underwater explosions in 2000 associated with the loss of a Russian submarine in the Barents Sea; the series of nuclear explosions carried out by India and Pakistan in 1998; and the nuclear tests conducted by North Korea in 2006 and 2009.

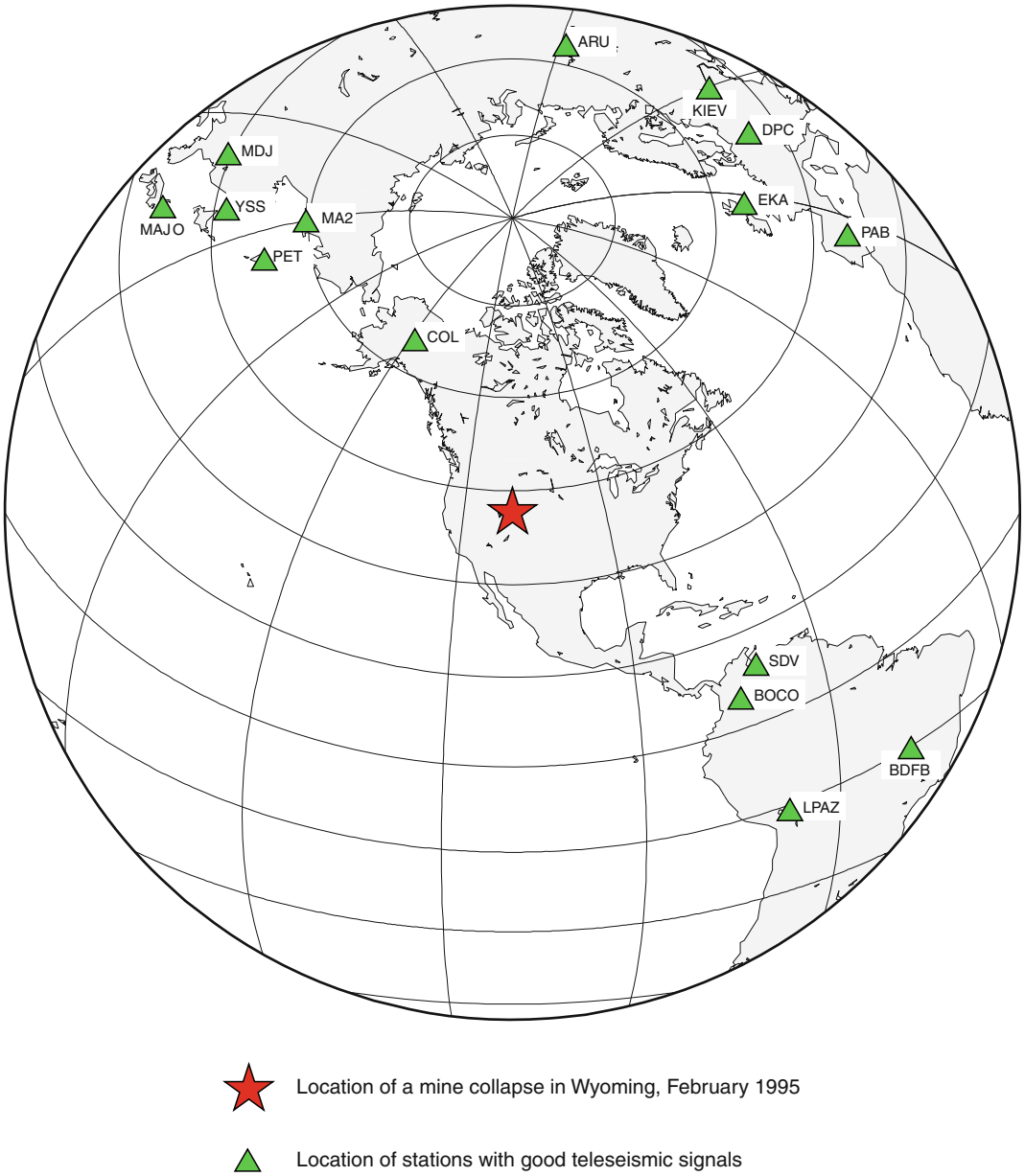
The mining collapses were seismically detected all over the world. For example, stations that detected the Wyoming event of 1995 are indicated in Figure 6. Mining collapses such as these have caused concern because their mix of surface waves and body waves as recorded teleseismically can appear explosion like using the classical M_s: m_b discriminant, as shown in Figure 2. But a careful analysis of regional and teleseismic waves from these events has showed that although the surface waves were quite weak, and in this respect seemed explosion like, they had the wrong sign. Therefore the motion at the source was *implosive* (the ground had moved inward toward the source) rather than *explosive*. Indeed, mining collapses are an implosion phenomenon, and it was important to learn that their implosive nature could be reliably determined from seismic recordings. Teleseismic waveforms from the Wyoming mine collapse are shown in Figure 7. This is an example of the use of what seismologists call the "first motion" of the *P*-wave, which is clearly downward in these data.

The Kara Sea earthquake was too small to apply the M_s: m_b discriminant (the surface waves were too small to measure reliably). This event showed the importance of accurate locations, and of using spectral ratios of regionally recorded *P*-waves and *S*-waves to discriminate small events (Richards and Kim, 1997).

As we have discussed earlier, the North Korea nuclear test of 2006 was of interest as an example of a nuclear explosion that was promptly detected globally, though its yield has been estimated at less than 1 kt. This event required regional seismic data in order to determine that indeed an explosion had been carried out and that the signals were not from an earthquake. Subsequently, xenon radionuclides were detected that decisively identified the explosion as nuclear.

Evasion

Several methods have been proposed, by which underground explosions might be concealed. One method is simply to make them small enough; but then there would be relatively little to learn, from the point of view of a weapons designer. The more important methods are those which combine as many features as possible, designed to reduce seismic signal-to-noise ratios at all relevant monitoring stations. Proposed methods include: emplacement of the nuclear device in material such as dry alluvium, to reduce the coupling of explosion energy into seismic signal (but that method is likely to result in leakage of detectable radioactivity); waiting until a sufficiently large natural earthquake occurs fairly near

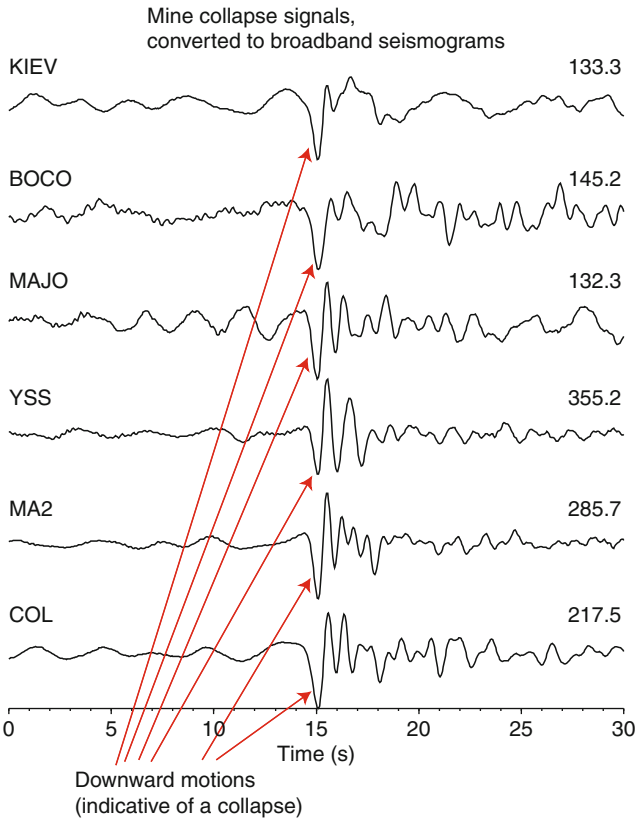


Seismic Monitoring of Nuclear Explosions, Figure 6 A global map showing stations recording teleseismic *P*-waves from a mine collapse in Wyoming. Its body wave magnitude was 5.3, and surface wave magnitude was 3.7. This combination is explosion like, as shown in Figure 2. (From Bowers and Walter, 2002.)

a test site (which presents the formidable challenge of identifying the event within a few minutes of its occurrence as large enough, and then within a couple of minutes executing the weapons test so that its seismic signals would hopefully be swamped by the large and prolonged signals from the earthquake); and setting off a sequence of several explosions that are designed to simulate a natural earthquake signal.

Careful study of each of these methods indicates that they are relatively ineffective in comparison with the methods known as cavity decoupling and mine masking, which we next discuss, and which are widely regarded as setting the practical levels down to which seismic monitoring of nuclear explosions is possible.

When an underground explosive device is tightly packed into its hole ("tamped" or "fully coupled"), and is detonated at sufficient depth to contain all radioactive products, a shock wave travels some distance from the shot-point out into the surrounding rock at speeds that exceed the normal *P*-wave speed. This nonlinear phenomenon reduces at sufficient distance from the shot-point, and thereafter the wave propagation can be regarded as



Seismic Monitoring of Nuclear Explosions, Figure 7 The *P*-wave signals of the Wyoming mine collapse are shown at six teleseismic stations, processed to bring out the fact that the first motion of the ground at these stations is downward, indicative of an implosion rather than explosion (for which the first motion would be upward). (From Bowers and Walter, 2002.)

elastic. The so-called "elastic radius" for a tamped explosion, i.e., the radius beyond which wave propagation is linear, is roughly 100 meters times the cube root of the yield (in kilotons).

If the explosion is set off inside a large underground cavity instead of being tamped, then the shock wave set up in the rock can be weakened or even eliminated, in which case only elastic waves are radiated. The explosion is said to be fully decoupled if only elastic waves result, and theoretical work reported in 1959 has addressed the question of how much weaker the seismic signal might be made. Theoretical work has indicated that signals could thereby be reduced by factors in the range 50–100, compared to a tamped explosion. The cavity radius itself is the "elastic radius" for a fully decoupled shot. For salt, the cavity radius required for full decoupling has been estimated at about 25 m times the cube root of the yield (in kilotons). For hard rock, the cavity size for full decoupling is comparable; for weak salt it is somewhat greater. See Sykes (1996) for further discussion, and Denny and Goodman (1990) for estimates of the decoupling factor derived from the practical experience in 1966 of carrying out a small nuclear explosion (about 0.38 kt) in the cavity produced by a tamped shot of 5.3 kt conducted 2 years earlier in a Mississippi salt dome. They conclude that the amplitude reduction is about 70, at low frequencies, for salt. At frequencies that have conventionally been used for seismic monitoring, the seismic signal strength is proportional (very roughly) to the volume within the elastic radius. This volume is substantially reduced by fully decoupling, which is the reason why cavity decoupling has been proposed as offering the technical possibility of a clandestine program of nuclear testing. However, the signal strength is not nearly so strongly reduced, by decoupling, at frequencies above that associated with resonances of the internal surface at the elastic radius. In practice, the frequency above which decoupling is likely to be substantially less effective is around 10–20 Hz, divided by the cube root of the yield (in kilotons). The overall effect on the seismic signals from a fully decoupled shot of yield Y, given the results of Denny and Goodman, is to make these signals look like those from a tamped shot of yield $Y/70$.

A thorough discussion of decoupling as an evasion scenario would have to include several non-seismological considerations. These include: the military significance of being able to carry out nuclear tests up to various different yield levels (e.g., 0.1, 1, or 10 kt.); and the political consequences if a clandestine test program were uncovered. Technical considerations include methods of (clandestine) cavity construction, and the capabilities of non-seismological surveillance techniques. Leakage of radioactivity from an underground cavity would be a challenge, given that much of the energy of a decoupled explosion goes into pumping up the pressure in the cavity. While some assert that clandestine use of cavity decoupling would be so difficult to execute that it belongs to the realm of fantasy, others have been persuaded that the risk might indeed be manageable, and that estimates of concealable yields, under this evasion scenario, must be made. The NAS (2002) report describes ten "layers of difficulty" with successfully hiding an underground nuclear explosion via cavity decoupling, concluding that even a nation committing significant resources to this work could not have confidence in being able to get away with tests above 1 or 2 kt.

The evasion scenario known as mine masking hypothesizes the execution of an underground nuclear weapon test explosion in a mining region, concurrently with a large mine blast. Such blasts in a big commercial operation consist of hundreds of separate charges, fired in sequence to break and/or move as much rock as possible, in a procedure known as ripple-firing (Khalturin et al., 1998). Regardless of the logistical difficulties of such a scenario, estimates of the possibilities of concealment via this approach can come from taking examples of signals from large mine blasts, and signals from small underground nuclear explosions, then adding them together before subjecting them to the methods used to discriminate between various types of seismic events. What is

typically found, is that the maximum size of the identifiable waves (e.g., the *P*-waves) from the mine blast is about that expected from individual sub-blasts (commonly called "delays"), and these amplitudes are spread out over a longer time in seismograms.

A study of mine masking possibilities by Smith (1993) used several different examples of mine-blast seismograms together with single-fired explosion records, and found a number of features that could be used to identify a simultaneous shot within a ripple-fired blast. He concluded that to conceal a single-fired deep detonation (depth is required for containment of radionuclides), the single explosive shot should not exceed 10% of the total explosive.

The conclusion here is that mine blasts are not effective for concealing large releases of energy at the level associated with kiloton-scale nuclear weapons tests, unless the nuclear explosion were subject to efforts at decoupling. Again non-seismic considerations arise, including an assessment of the plausibility of carrying out a complicated decoupled and masked nuclear explosion at the same time and location as a large mine blast that would itself attract some level of monitoring attention – particularly if the seismic signals seemed unusual in comparison with those from prior blasting in the region.

Event detection capability of the international monitoring system

In 1976, a group of international scientists was established at the Conference on Disarmament in Geneva, for the study of monitoring technologies and data analysis methods in the context of supporting a future test ban treaty. This group of scientific experts (GSE) played an essential role in laying the scientific groundwork for the final stage of CTBT negotiations conducted from 1994 to 1996. Prior to the negotiation, GSE organized a series of technical tests – GSETT-1 in 1984, GSETT-2 in 1991, and GSETT-3 in 1995. These tests contributed significantly to the development of the international system being built today to support treaty verification.

The finalized sections of the CTBT include an extensive description of networks to monitor treaty compliance using hydroacoustic, infrasound, and radionuclide technologies as well as seismological methods. The CTBT Organization (CTBTO) operates an International Monitoring System specified in treaty text, as well as an International Data Centre to analyze signals sent via satellite to headquarters in Vienna. Extensive descriptive material on these networks is available online (see http://www.ctbto.org).

To implement the CTBT seismic monitoring system, a sequential four-step process is needed to build each station (CTBTO PrepComm, 2009): (1) Site survey, (2) Installation, (3) Certification, and (4) Operation. It must be demonstrated for IMS stations that data received at the International Data Centre (IDC) are *authentic*. This is achieved through a special digital "signature" embedded in the data flow from each station. The IMS station must be certified to ensure that all of its equipment, infrastructure, and settings meet the technical specifications set by the CTBTO, and to also ensure that all data are transmitted to the IDC through the Global Communication Infrastructure (GCI) in a timely manner.

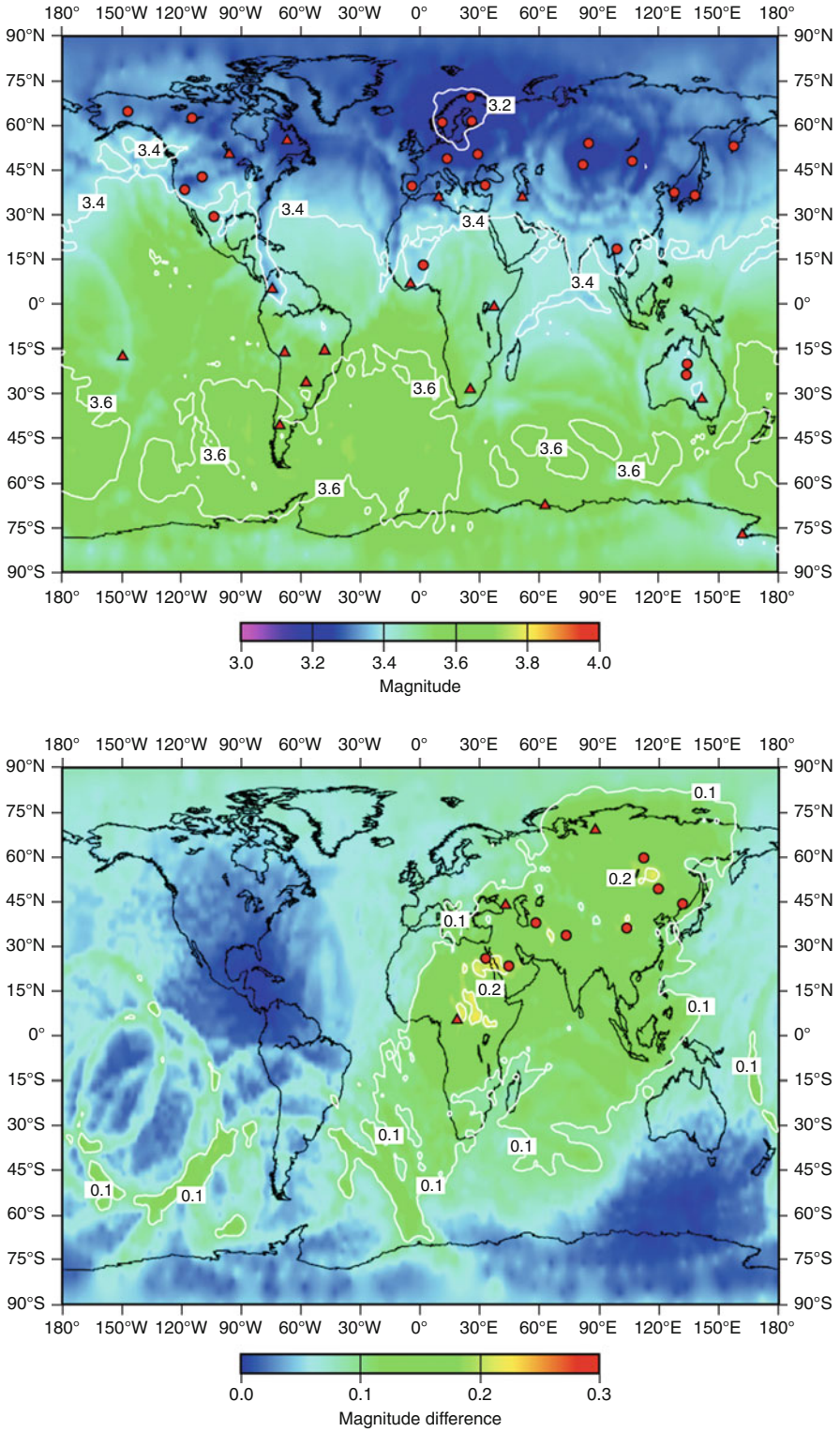
Here, we note that the primary seismographic network is to consist of 50 stations, many of them arrays; and that location estimates are based upon detection of signal at 3 stations or more. An auxiliary network of 120 continuously operating stations is to provide seismic waveform data, again via satellite, in order to help characterize the events detected by the primary network. Although these two networks are not completely built (as of 2010), there are enough stations operating to provide good indications of what the detection capability will be when all stations are installed and providing data.

Figure 8 shows maps of the detection capability of the primary seismic network of the IMS. The upper figure shows the actual capability of 38 operating stations based upon experience in the year 2007. The lower figure shows how much this capability is expected to improve when 11 additional stations are operational, most of them in Eurasia. Capability is expressed in terms of magnitude thresholds, above which 90% of the seismic events are expected to be detected at enough stations to provide a location estimate. The work of identifying events is left to member states. This work is not just a technical matter since it is a political act for one country to make an allegation that another country has committed a treaty violation. The evidence in support of such an allegation can come from the IMS and IDC, as well as from the National Technical Means of member states, and/or from a subset of the thousands of seismographic stations operated around the world for purposes not directly related to monitoring for nuclear explosions.

Summary

We have described the basic steps in monitoring nuclear explosions, and have emphasized the seismic monitoring system specified by the Comprehensive Nuclear Test Ban Treaty of 1996.

When the treaty was being negotiated, the goal for the International Monitoring System was that it be capable of detecting and identifying treaty violations – nuclear explosive tests – at the 1 kt level and higher, if they were not evasively tested. Recognizing that a 1 kt underground nuclear explosion has a magnitude in the range about 4–4.5, if it is conducted in the way that almost all the more than 1,500 prior underground nuclear explosions were carried out (i.e., well tamped and not with intent to reduce the signals picked up by monitoring networks), the evidence from Figure 8 is that this design capability has been significantly exceeded. For almost the entire northern hemisphere, including Eurasia and North America, capability is good down to about magnitude 3.3. This corresponds to a yield of less than 100 t (0.1 kt) for a well-tamped explosion in hard rock. Only time will tell whether this capability, combined with other monitoring assets, is deemed adequate to support entry into force of the CTBT.



Seismic Monitoring of Nuclear Explosions, Figure 8 Maps showing the detection capability of the IMS primary seismographic network. The upper figure shows the capability of the network in late 2007, with 38 stations sending data to the IDC. The capability is represented by the magnitude of the smallest seismic event that would be detected with a 90% probability by three stations or more. The lower figure shows the estimated improvement over this capability that could be achieved by bringing 11 of the remaining 12 primary seismic stations into operation. (From Kværna and Ringdal, 2009.)

Acronyms

CTBT–Comprehensive Test Ban Treaty or Comprehensive Nuclear-Test-Ban Treaty (its formal name)
CTBTO–CTBT Organization
IDC–International Data Centre (of the CTBTO)
IMS–International Monitoring System (of the CTBTO)
LTBT–Limited Test Ban Treaty
TTBT–Threshold Test Ban Treaty

Bibliography

Bennett, T. J., Oancea, V., Barker, B. W., Kung, Y.-L., Bahavar, M., Kohl, B. C., Murphy, J. R., and Bondár, I. K., 2010. The Nuclear Explosion Database (NEDB): a new database and web site for accessing nuclear explosion source information and waveforms. *Seismological Research Letters*, **81**, 12–25.

Bowers, D., and Walter, W. R., 2002. Discriminating between large mine collapses and explosions using teleseismic P Waves. *Pure and Applied Geophysics*, **159**, 803–830.

Bowers, D., and Selby, N. D., 2009. Forensic Seismology and the Comprehensive Nuclear-Test-Ban Treaty. *Annual Reviews of Earth and Planetary Sciences*, **37**, 209–236.

CTBTO PrepComm, 2009. Verification regime (available online at http://www.ctbto.org/verification-regime/).

Dahlman, O., Mykkeltveit, S., and Haak, H., 2009. *Nuclear Test Ban: Converting Political Visions to Reality*. Berlin: Springer.

Denny, M. D., and Goodman, M. D., 1990. A case study of the seismic source function: Salmon and Sterling reevaluated. *Journal of Geophysical Research*, **95**, 19705–19723.

Der, Z., Mcelfresh, T., Wagner, R., and Burnetti, J., 1985. Spectral characteristics of *P* waves from nuclear explosions and yield estimation. *Bulletin of the Seismological Society of America*, **75**, 379–390 (also erratum, **75**, 1222–1223).

Ford, S. R., Dreger, D. S., and Walter, W. R., 2009. Identifying isotropic events using a regional moment tensor inversion. *Journal of Geophysical Research*, **114**, B01306, doi:10.1029/2008JB005743.

Khalturin, V. I., Rautian, T. G., and Richards, P. G., 1998. The seismic signal strength of chemical explosions. *Bulletin of the Seismological Society of America*, **88**, 1511–1524.

Khalturin, V. I., Rautian, T. G., Richards, P. G., and Leith, W. S., 2005. A review of nuclear testing by the Soviet Union at Novaya Zemlya, 1955-1990. *Science & Global Security*, **13**, 1–42.

Kim, W.-Y., Simpson, D. W., and Richards, P. G., 1993. Discrimination of earthquakes and explosions in the Eastern United States using regional high-frequency data. *Geophysical Research Letters*, **20**, 1507–1510.

Koper, K. D., Herrmann, R. B., and Benz, H. M., 2008. Overview of open seismic data from the North Korean event of 9 October 2006. *Seismological Review Letters*, **79**, 178–185.

Kværna, T., and Ringdal, F., 2009. *Analysis of the IDC Reviewed Event Bulletin for Detection Capability Estimation of the IMS Primary and Auxiliary Seismic Stations*. Poster presented at the International Scientific Studies conference, June 10–12, 2009, Vienna, Austria.

Murphy, J. R., 1996. Types of seismic events and their source descriptions. In Husebye, E. S., and Dainty, A. M. (eds.), *Monitoring a Comprehensive Nuclear Test Ban Treaty. NATO ASI Series E*. Dordrecht: Kluwer Academic, Vol. 303, pp. 225–245.

Murphy, J. R., Rodi, W., Johnson, M., Sultanov, D. D., Bennett, T. J., Toksöz, M. N., Ovtchinnikov, V., Barker, V. W., Reiter, D. T., Rosca, A. C., and Shchukin, Y., 2005. Calibration of International Monitoring System (IMS) stations in Central and Eastern Asia for improved seismic event location. *Bulletin of the Seismological Society of America*, **95**, 1535–1560.

Myers, S. C., Begnaud, M. L., Ballard, S., Pasyanos, M. E., Phillips, W. S., Ramirez, A. L., Antolik, M. S., Hutchenson, K. D., Dwyer, J. J., Rowe, C. A., and Wagner, G. S., 2010. A crust and upper mantle model of Eurasia and North Africa for *Pn* travel time calculation. *Bulletin of the Seismological Society of America*, **100**, 640–656.

NAS, 2002. National Academy of Sciences report. *Technical Issues Related to the Comprehensive Nuclear-Test-Ban Treaty*, Washington, DC: National Academy Press.

Nuttli, O. W., 1986. Yield estimates of Nevada Test Site Explosions obtained from Seismic *Lg* Waves. *Journal of Geophysical Research*, **91**, 2137–2151.

OTA, 1988. Office of Technology Assessment, Congress of the United States, *Seismic Verification of Nuclear Testing Treaties*, OTA-ISC-361. Washington: U.S. Government Printing Office.

Richards, P. G., 1988. Seismic methods for verifying test ban treaties. In Schroeer, D., and Hafemeister, D. (eds.) Chapter 4 of *Nuclear Arms Technologies in the 1990's*. *AIP Conference Proceedings*. New York: American Institute of Physics, Vol. 178, pp. 54–108.

Richards, P. G., and Kim, W.-Y., 1997. Testing the nuclear test-ban treaty. *Nature*, **389**, 781–782.

Richards, P. G., Waldhauser, F., Schaff, D., and Kim, W.-Y., 2006. The applicability of modern methods of earthquake location. *Pure and Applied Geophysics*, **163**, 351–372.

Ringdal, F., 1985. Study of magnitudes, seismicity and earthquake detectability using a global network. In Kerr, A. U. (ed.) *The VELA Program: A Twenty-Five Year Review of Basic Research*. Defense Advanced Research Projects Agency, pp. 611–624.

Ringdal, F., Marshall, P. D., and Alewine, R. W., 1992. Seismic yield determination of soviet underground nuclear explosion at the shagan river test site. *Geophysical Journal International*, **109**, 65–77.

Smith, A. T., 1993. Discrimination of explosions from simultaneous mining blasts. *Bulletin of the Seismological Society of America*, **83**, 160–179.

Sykes, L. R., 1996. Dealing with decoupled nuclear explosions under a comprehensive test ban treaty. In Husebye, E. S., and Dainty, A. M. (eds.), *Monitoring a comprehensive nuclear test ban treaty*. Dordrecht: Kluwer Academic. NATO ASI series E, Vol. 303, pp. 247–293.

Timerbaev, R., undated. *On the "Threshold" Test Ban Treaties of 1974–76*. Paper available from the PIR Center as http://www.pircenter.org/kosdata/page_doc/p1650_1.pdf

Waldhauser, F., and Schaff, D. P., 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods. *Journal of Geophysical research*, **113**, doi:10.1029/2007JB005479.

Yang, X., North, R., Romney, C., and Richards, P. G., 2003. Worldwide nuclear explosions. In chapter 84 of *International Handbook of Earthquake and Engineering Seismology*, Part B, edited by Lee, W. H. K., Kanamori, H., Jennings, P., and Kisslinger, C. on behalf of the International Association of Seismology and Physics of the Earth's Interior, Academic, pp. 1595–1599.

Cross-references

Body Waves
Earthquake, Magnitude
Seismic Instrumentation
Seismological Networks
Seismology, Monitoring of CTBT
Surface Waves

SEISMIC NOISE

Dhananjay Kumar[1], Imtiaz Ahmed[2]
[1]Chevron, Houston, TX, USA
[2]BP, Houston, TX, USA

Definition

Seismic noise. Noise is the undesirable part of seismic data that is not signal, and signal is what fits our conceptual model.
Multiple. A seismic event that experiences more than one reflection in the subsurface.
SNR. Signal-to-noise ratio – is a measure of strength of signal compared to noise, and it is a measure of seismic data quality.
Active seismic. Seismic data recordings of artificial man-made sources such as dynamite and vibroseis in land surveys and air gun in marine surveys. Example: exploration and engineering seismology.
Passive seismic. Seismic data recordings of natural sources such as earthquake, solar waves, and ocean waves. Example: earthquake seismology.

Introduction

Seismic noise comprises all of the unwanted recorded energy that contaminates seismic data. A part of the seismic energy is considered noise if it does not fit the conceptual model of seismic signal. Seismic noise can be random or coherent. The identification of seismic noise depends on the type of data analysis and the type of data available – a part of data treated as noise in one application can be signal in another application. For example, the S-wave energy is generally considered noise in a P-wave processing project; vice versa, the P-wave energy is considered noise in an S-wave processing project.

Historically, only the traveltime information of the seismic data was used to study the subsurface. For example, the knowledge about the earth's deep interior was primarily derived from seismic traveltime information recorded during deep earthquakes. Also, only the traveltime information was used to derive the structural image of subsurface for exploration projects. As seismic technology has advanced and the appetite for understanding complicated geological features has increased, there has been a push toward technologies using the complete waveforms (amplitude and traveltime) in seismic analysis. Full waveform inversion technology (Tarantola, 1986) is one such example that uses the complete waveform to estimate properties of the subsurface from the seismic data. The success of such technologies is understandably very dependent on clean, noise-free seismic data.

It is essential that seismic data are carefully processed to derive high quality seismic images (see Yilmaz, 2001 for seismic data processing). One of the major challenges in seismic data processing is to separate noise from signal or to attenuate noise. In practice, noise cannot be completely attenuated and occasionally it is not even desired to attenuate noise but to use it as signal. The objective of the noise attenuation or noise separation process in seismic data processing is to enhance signal-to-noise ratio (SNR). There have been significant progress in data processing to improve SNR; advances have been made in random noise attenuation (Yilmaz, 2001) and coherent noise attenuation (see Weglein and Dragoset, 2005 for multiple attenuation methods).

Recently, there have been various efforts to use seismic noise as signal, for example: (1) using multiple reflected energy (multiples) in seismic migration and inversion to image subsurface; and, (2) using very low and very high frequency passive seismic signal for reservoir monitoring. In the following sections, we write brief descriptions about the types of seismic noise, the noise attenuation techniques, and how seismic noise can be useful.

Types of seismic noise

There are two types of seismic noise: random noise and coherent noise. In a multichannel seismic dataset, random noise does not correlate either with the neighboring channels (i.e., no spatial correlation) or along the same channel (i.e., no temporal correlation). Coherent noise, however, is part of the data that correlates spatially and/or temporally. Random noise is easier to attenuate during seismic data processing. Coherent noise is difficult to attenuate in processing; therefore, residual coherent noise can interfere with real signal and be misinterpreted as signal. The possible sources of these seismic noises can be placed under four categories: (1) ambient sources, (2) wave propagation related noise, (3) data acquisition related noise, and (4) data processing artifacts. The severity and types of noise can differ between marine and land acquisition environment.

Ambient noise is the noise from unwanted sources like wind, swell, power line, activities on nearby road, surface facility, marine activities like ships and marine animals, and other cultural noise. Ambient noise can be present in various forms on seismic data, such as linear features, localized very high amplitude response, and mono-frequency events. Ambient noise can be random noise and coherent noise.

Wave propagation related noise includes the surface waves, multiples, and geologic noise. Seismic response for an active source survey include primary reflection event (e.g., incident P-wave reflected back as P-wave), refraction, ground roll, mode converted event (e.g., incident P-wave reflected as S-wave), and several events reflected multiple times in the subsurface (multiples). In reflection seismology, we are interested in P-wave (and/or S-wave) primary reflections. All the other coherent energies recorded are considered noise. Multiples are considered the major source of coherent noise in the seismic experiment and are really difficult to attenuate. Wave propagation related noise also includes seismic response from unwanted geology (complicated stratigraphy,

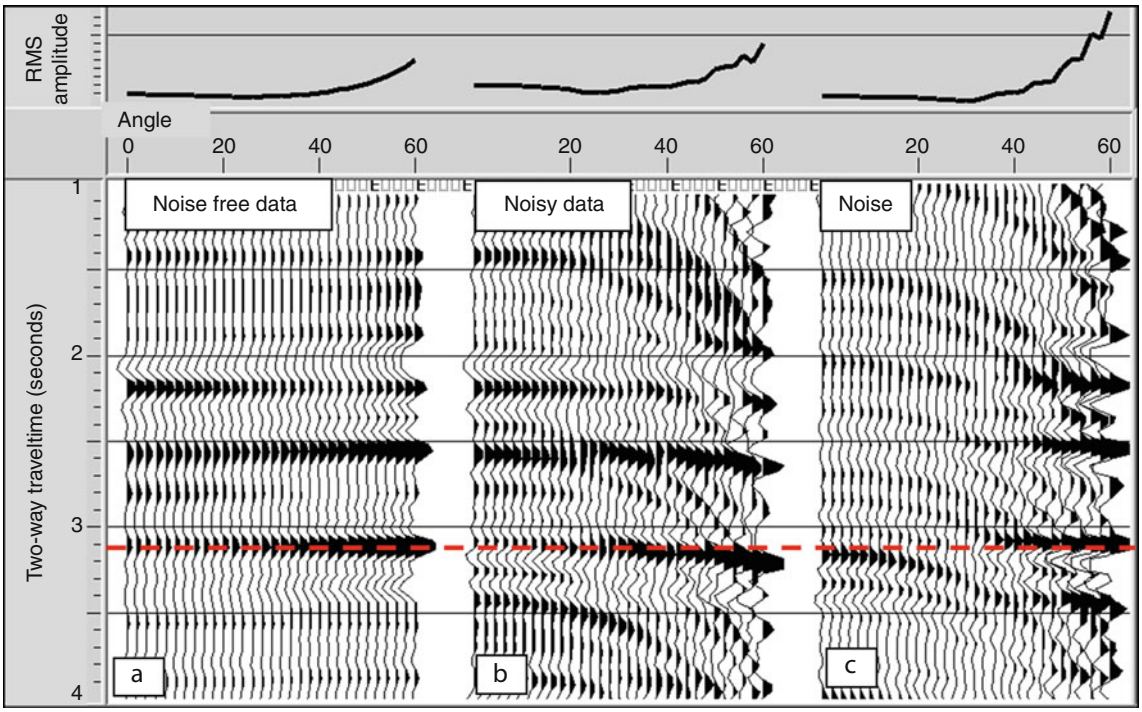
shallow gas, and faults) not considered in seismic analysis; they are known as geologic noise.

Data acquisition related noise is due to poor source and receiver coupling to the earth, source and recording instruments-generated noise, and acquisition footprint related to the acquisition design. Noise from poor coupling and noise related to instruments can be easily identified and attenuated. Acquisition footprint related to the acquisition design, observed as linear trend in seismic amplitude map view, is common in seismic data and can be suppressed in processing.

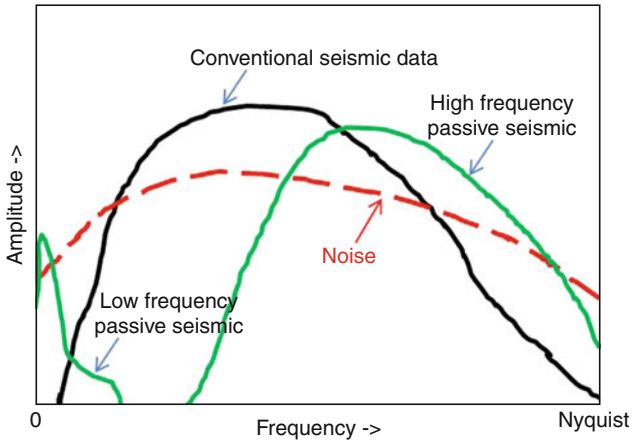
Seismic data processing is another source of noise in processed seismic data. There are various factors affecting seismic reflection amplitude that do not contain subsurface information (Sheriff, 1975), but it is impossible to correct for all the factors affecting amplitudes. Many approximations and assumptions are made in seismic data processing depending on computer resources availability, project timeline, understanding of the physics of wave propagation, and the type of seismic data available. Some of these factors may alter the data and may introduce noise in the recorded data. Some examples are: (1) noise introduced in seismic data due to poor multiple attenuation and poor normal move-out (NMO) correction (Figure 1); (2) noisy subsurface image due to imperfect velocity model used and/or approximate physics used for migration; (3) noise introduced due to inaccurate amplitude calibration of the raw amplitudes from the processed seismic data with synthetics seismic amplitude for quantitative AVO/AVA (amplitude variation with offset/angle) analysis; (4) artifacts introduced from the process of frequency enhancement to broaden frequency bandwidth for better depth resolution; and (5) artifacts introduced from the process of interpolation and regularization to compensate for the irregular acquisition geometry.

Enhancing signal over noise

There is always some noise present in the seismic data. Figure 2 schematically shows amplitude spectra of various seismic signals and noise, and noise is present in the whole signal bandwidth. Thus, the objective is to enhance SNR by noise attenuation or separation so that data can be effectively used for analysis. An important step in noise



Seismic Noise, Figure 1 Example of seismic noise. Shown in this figure is normal move-out (NMO) corrected synthetic seismic angle gather for AVA analysis with angle (in degrees) as *x*-axis and two-way traveltime (in seconds) as *y*-axis for: noise-free case (**a**), with noise (**b**), and the noise present in data (**c**). The response in the third panel (**c**) is the difference between response in the first panel (**a**) and the second panel (**b**). On the top of angle gather is the root mean square (RMS) amplitude value as a function of angle; the RMS value is computed in entire time gate shown here. In (c), the dipping coherent noises are multiples and the strong amplitudes at far angles are residual data after NMO. Thus, part of the data became residual data (un-flattened data after NMO in (b) is noise in AVA analysis) due to imperfect seismic velocity and/or seismic modeling method used in NMO (see red lines marked at 3,100 ms across three gathers for residual data).



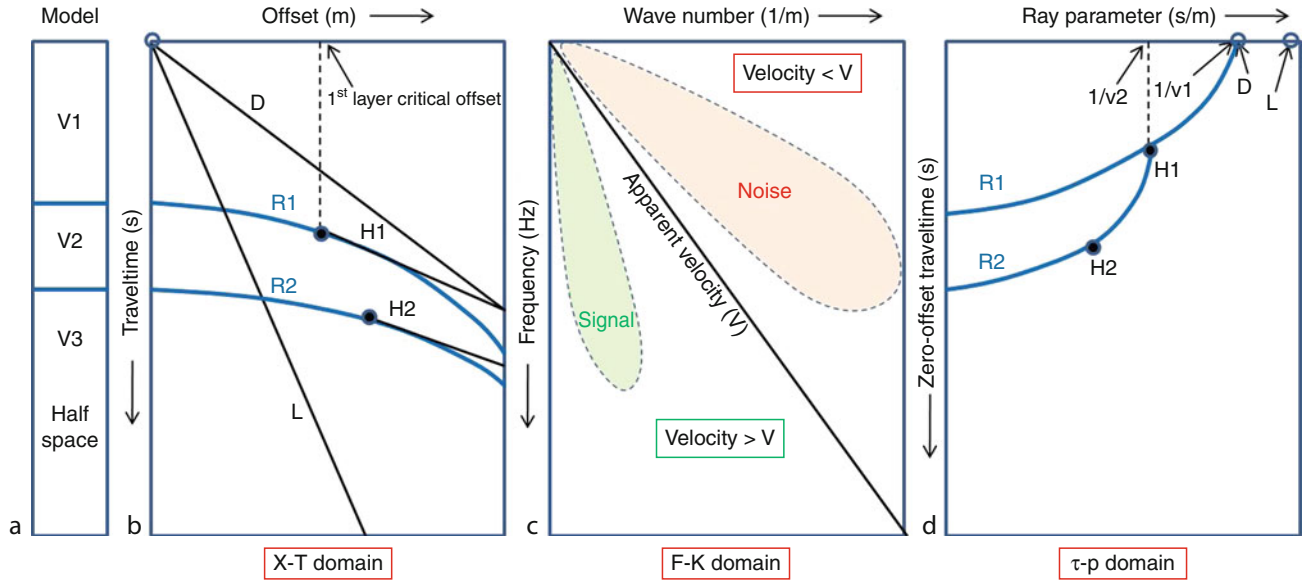
Seismic Noise, Figure 2 Schematic amplitude spectrum: for conventional broadband seismic data (active seismic for exploration range 10–60 Hz, and passive seismic for earthquake range from 10–100 Hz), low frequency passive seismic data (typically less than 10 Hz), high frequency passive seismic (microseismic) data (typically 30–300 Hz), and noise. Because noise is present at all frequencies, there is a need to do noise attenuation so that we get signal-to-noise ratio greater than 1. Spectrum is plotted for near zero frequency to Nyquist frequency. Nyquist frequency is the highest frequency without aliasing; beyond this frequency signal cannot be reliably estimated. Note that passive low and high frequency data can be treated as noise in broadband seismic data.

attenuation/separation is to identify signal from noise. White (1984) discusses spectral methods for signal and noise estimation on seismic reflection data. Kanasewich (1990) and Chapter 6 of Yilmaz (2001) have good discussion on noise and multiple attenuation methods with data examples. Noise level in data compared to signal can be estimated as the ratio of autocorrelation of data to a cross correlation of data. This is because autocorrelation of data represents both signal and noise but cross correlation of data with a representative signal trace (can be local partial stack of data) represents signal in data. Signal and noise separation process can be broadly divided into two methods: prediction based methods and separation based methods. By performing seismic simulation over a known (conceptual) model and comparing synthetics to the field seismic data, signal and noise can be identified on data, and therefore noise can be attenuated. Also, signal and noise representation differs in different data domains. Therefore, a suitable data domain can be identified for the optimum separation of the two. Figure 3 schematically shows various seismic events for a three-layer model in native acquisition (X-T) domain and transformed domains: (1) X-T domain (represents spatial-time domain), (2) F-K domain (represents frequency – wave number domain), and (3) τ-p domain (representing zero-offset traveltime – ray parameter domain). The key to successful noise attenuation is large separation between signal and noise in a specific data domain. Different types of noise may require different domains for better separation between signal and noise. For example, a linear event on a shot gather (X-T domain) maps to a radial line in the F-K domain and it maps onto a point in the τ-p domain, and thus can be rejected by F-K filtering and τ-p filtering, respectively. Caution should be observed while performing the forward and inverse transforms, as some implementations may not be completely reversible and may also introduce artifacts.

Random noise is not correlated and can be attenuated easily during data processing. One of the most robust methods to attenuate random noise on multichannel data by a factor of square root of N is by stacking (compositing) data from N channels. If prestack data is needed, however, F-X deconvolution (predictive deconvolution) is effective in random noise attenuation, where F-X corresponds to frequency – space domain. Deconvolution operation (Yilmaz, 2001) is designed in space domain to predict coherent signals, and then coherent signals can be subtracted from the seismic record.

Coherent noise is relatively more difficult to attenuate. An appropriate data domain (Figure 3) can be selected to distinguish signal from noise using the noise characteristics, like velocity and frequency. Radon filtering in τ-p domain and wavelet transform filtering in F-K domain is very effective method for coherent noise attenuation. In many field data cases, signal and noise are not well separated in any data domain and therefore it is very difficult to attenuate the noise. Recently a version of F-K filtering called Curvelet transform filtering has shown better results for both random and coherent noise attenuation (Neelamani et al., 2008). Multiples are example of coherent noise present in the data. There has been extensive research in the field of multiples suppression. The surface related multiple elimination (SRME) has been heavily relied upon to suppress multiples (Berkhout and Verschuur, 1997). This technique is based on the stationary phase theory of the seismic signal and the multiples are predicted by convolving traces with one another and then stacking these convolved traces.

Another way to have better SNR in seismic data is via improved seismic acquisition techniques. There have been various developments in seismic acquisition on land and in marine environment. For example, over/under marine seismic acquisition promises better SNR in both low and high frequency range (Moldoveanu et al., 2007). In over/under towed-streamer acquisition pairs of streamer are deployed at two different depths in the same vertical plane. Other advances in seismic acquisition promise cleaner and better seismic images, for example, Multicomponent Ocean Bottom recording, Wide Azimuth recording, and Simultaneous Sources recording. The multiples suppression is strongly dependent on the acquisition effort. For example, the suppression of multiples in the case of Wide Azimuth acquisition is the result of a natural weighting of the traces going into the stack



Seismic Noise, Figure 3 Seismic responses in various data domains for a three-layer model. The time domain earth model is shown in left (**a**) with seismic velocity for three layers as V1, V2, and V3. The second from left is the seismic response for a three-layer model in X-T domain (**b**), where D represents direct arrival, L represents linear noise, such as ground rolls, R1 and R2 are primary reflection events from 1st and 2nd layer interfaces, and H1 and H2 are refraction events. The third from left shows the seismic response in F-K domain (**c**), where noise (D, L events) is on the upper right corner with apparent velocity less than the reference apparent velocity (V) and signal (R1, R2, H1, H2 events) are on the lower left corner corresponding to lower wave numbers. The right most plot (**d**) shows the seismic response in the τ-p domain, where the linear event in X-T domain plots on a point in the τ-p domain (Diebold and Stoffa, 1981).

because of the areal nature of the acquisition (VerWest and Lin, 2007).

Some noise will always remain in the seismic data even after careful seismic processing. Also, there are some noises that are still not understandable and difficult to attenuate. Therefore, it is important to incorporate noise in the seismic data analysis – for example, in seismic inversion the data should not be over fitted as it might be fitting the noise.

Use of noise as signal

There have been significant efforts recently to use all of the recorded energy in the seismic data. Traditionally, only the primary reflection wavefield is used in reflection seismic imaging, while the multiple/ghost is discarded. However, the ghost/multiples can be used as signal, for example, as an input to mirror migration (Verm, 1987) to produce superior shallow images for the sparse receiver ocean bottom data where illumination holes deteriorate the primary images (Clarke et al., 2006; Dash et al., 2009). Note that multiples are also generated from the same source as primary, they travel longer paths and contain more information than primary, and therefore in some circumstances multiples can be more useful than primaries. Full waveform inversion is another technique that uses both the primary and the multiples to invert for the subsurface parameters more effectively.

Multiples have also been used to interpolate for missing near offsets in seismic recording using the technique of interferometry (Schuster, 2009). Seismic interferometry can also be used to extract primary signal from background noise and multiples by simply cross-correlating recorded data at two different stations (Sabra et al., 2005; Curtis et al., 2006; Draganov et al., 2007).

Some new seismic acquisition methods have been useful in using conventional multiples to improve SNR and/or extract primary. For example, over/under towed-streamer acquisition technology uses multiples to improve seismic data quality (Moldoveanu et al., 2007), and there is a possibility to estimate primaries from surface multiples from data recorded with simultaneous sources also called blended data (van Groenestijn and Verschuur, 2010). In the simultaneous source acquisition (Berkhout, 2008; Hampson et al., 2008) multiple sources are fired in a short time interval to speed up the acquisition.

Converted S-wave data, regarded as noise in reflected P-wave data imaging, has been quite successful in imaging gas reservoirs and areas where we have shallow gas anomalies (Tatham and McCormack, 1991). The property that S-wave does not get as attenuated as P-wave when the wavefield travels through these porous medium helps create better images through pure S-wave or converted S-wave imaging.

Low frequency earth's ambient passive response caused by natural phenomenon (such as wind, ocean waves, and human-made noise) and high frequency passive seismic response due to small earthquakes (microseisms) caused by induced fractures in a petroleum reservoirs are considered noise in a broadband active seismic data, but they can be effectively used to study subsurface. Low frequency earth's ambient noise can be used to extract signal (Draganov et al., 2007), and the low frequency passive seismic anomaly can be correlated with the presence of hydrocarbon (Saenger et al., 2009); however the research is still in its early stages. High frequency passive seismic data (also called microseismic data) are effectively used in hydraulic fracture monitoring (Warpinski, 2009) by locating microseisms induced by hydrocarbon production related activities.

Summary

Seismic noise is an integral part of the seismic record and is defined as all unwanted seismic energy on data. It can be divided into two categories: random and coherent noises. Random noise is not correlated among traces and is easier to attenuate compared to coherent noise that is spatially and/or temporally correlated. Multiples and geologic noise that are coherent noise are more difficult to attenuate and often interfere with seismic signal and makes seismic analysis challenging. Strategies for seismic noise attenuation are needed to preserve the seismic signal of interest and to improve signal-to-noise ratio (SNR). The success in noise attenuation lies in identification and then separation or prediction of signal and noise. Transformation of data to different data domains (X-T, F-K, τ-p, curvelet, wavelet domains) have helped in better separating noise from signal. Wave-equation extrapolations, inverse scattering methods, surface related multiple elimination, deconvolution, etc., model the noise and/or data in the process of noise attenuation. Advances in seismic data acquisition and processing have been made to improve SNR. Recently various efforts have been made to use noise as signal and are an active topic of research. This includes using multiples as well as primaries in seismic migration and inversion, and using low and high frequency passive seismic data in imaging subsurface.

Bibliography

Berkhout, A. J., 2008. Changing the mindset in seismic acquisition. *The Leading Edge*, **27**, 924–938.

Berkhout, A. J., and Verschuur, D. J., 1997. Estimation of multiple scattering by iterative inversion, Part I: theoretical considerations. *Geophysics*, **62**, 1586–1595.

Clarke, R., Xia, G., Kabir, N., Sirgue, L., and Mitchell, S., 2006. Case study: a large 3D wide azimuth ocean bottom node survey in deepwater GOM. In *76th Annual International Meeting, Society of Exploration Geophysicists*, Expanded Abstracts, 1128–1132.

Curtis, A., Gerstoft, P., Sato, H., Snieder, R., and Wapenaar, K., 2006. Seismic interferometry – turning noise into signal. *The Leading Edge*, **25**, 1082–1092.

Dash, R., Spence, G., Hyndman, R., Grion, S., Wang, Y., and Ronen, S., 2009. Wide-area imaging from OBS multiples. *Geophysics*, **74**(4), Q41–Q47.

Diebold, J., and Stoffa, P. L., 1981. The traveltime equation tau-p mapping and inversion of common midpoint data. *Geophysics*, **46**, 238–254.

Draganov, D., Wapenaar, K., Mulder, W., Singer, J., and Verdel, A., 2007. Retrieval of reflections from seismic background-noise measurements. *Geophysical Research Letters*, **34**, L04305, doi:10.1029/2006GL028735.

Hampson, G., Stefani, J., and Herkenhoff, F., 2008. Acquisition using simultaneous sources. *The Leading Edge*, **27**, 918–923.

Kanasewich, E. R., 1990. *Seismic Noise Attenuation. Handbook of Geophysical Exploration, Seismic Exploration*. New York: Pergamon, Vol. 7.

Moldoveanu, N., Combee, L., Egan, M., Hampson, G., Sudora, L., and Abriel, W., 2007. Over/under towed-streamer acquisition: a method to extend bandwidth to both higher and lower frequencies. *The Leading Edge*, **26**, 41–58.

Neelamani, R., Baumstein, A. I., Gillard, D. G., Hadidi, M. T., and Soroka, W., 2008. Coherent and random noise attenuation using the curvelet transformation. *The Leading Edge*, **27**, 240–248.

Sabra, K. G., Gerstoft, P., Roux, P., and Kuperman, W. A., 2005. Extracting time-domain Green's function estimates from ambient seismic noise. *Geophysical Research Letters*, **32**, L03310, doi:10.1029/2004GL021862.

Saenger, E. H., Schmalholz, S. M., Lambert, Marc-A, Nguyen, T. T., Torres, A., Metzger, S., Habiger, R. M., Muller, T., Rentsch, S., and Mendez-Hernandez, E., 2009. A passive seismic survey over a gas field: analysis of low-frequency anomalies. *Geophysics*, **74**(2), 029–040.

Schuster, G. T., 2009. *Seismic Interferometry.* UK: Cambridge University Press.

Sheriff, R. E., 1975. Factors affecting seismic amplitudes. *Geophysical Prospecting*, **23**, 125–138.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, **51**, 1893–1903.

Tatham, R. H., and McCormack, M. D., 1991. *Multicomponent seismology in petroleum exploration*. Tulsa: Society of Exploration Geophysicists.

van Groenestijn, G. J. A., and Verschuur, D. J., 2010. Using surface multiples to estimate primaries by sparse inversion from blended data. *Geophysical Prospecting*, doi:10.1111/j.1365-2478.2010.00894.x.

Verm, R., 1987. Imaging VSP's 3 kilometers beyond the borehole receiver. In *Offshore Technology Conference Proceedings*, Paper 5570.

VerWest, B. J., and Lin, D., 2007. Modeling the impact of wide-azimuth acquisition on subsalt imaging. *Geophysics*, **72**, 241–250.

Warpinski, N., 2009. Microseismic monitoring: inside and out. *Journal of Petroleum Technology*, **61**, 80–85.

Weglein, A. B., and Dragoset, W. H. (eds.), 2005. Multiple attenuation. *Geophysics Reprint Series No. 23, Society of Exploration Geophysicists*, Tulsa.

White, R. E., 1984. Signal and noise estimation from seismic reflection data using spectral coherence methods. *Proceedings of the IEEE*, **72**(10), 1340–1356.

Yilmaz, O., 2001. *Seismic data analysis: processing, inversion, and interpretation of seismic data*. Tulsa: Society of Exploration Geophysicists.

Cross-references

Earthquakes, Source Theory
Seismic Data Acquisition and Processing
Single and Multichannel Seismics

SEISMIC PHASE NAMES: IASPEI STANDARD

Dmitry A. Storchak[1], Johannes Schweitzer[2], Peter Bormann[3]
[1]International Seismological Centre (ISC), Thatcham, Berkshire, UK
[2]NORSAR, Kjeller, Norway
[3]Department 2: Physics of the Earth, GFZ German Research Center for Geosciences, Potsdam, Germany

Introduction

The working group (hereinafter WG) on the Standard Seismic Phase Names was set up by the IASPEI Commission on Seismological Observation and Interpretation (CoSOI) in 2001. The WG was chaired by D. A. Storchak, and included R. D. Adams, P. Bormann, E. R. Engdahl, J. Havskov, B. L. N. Kennett, and J. Schweitzer. The WG put together a modified standard nomenclature of seismic phases that was meant to be concise, consistent, and self-explanatory on the basis of agreed rules. The list was not meant to satisfy specific requirements of seismologists to name various phases used in a particular type of research. Instead, it was hoped that the new list would ensure an expanded standardized data reporting and exchange by data analysts and other users. After numerous consultations with the seismological community, the Standard Seismic Phase List was finalized and adopted by the CoSOI/IASPEI at its meeting in Sapporo on July 04, 2003.

The original list of standard seismic phase names was first published as part of the New Manual of Seismological Observatory Practice (Storchak et al., 2002) and then the version formally approved by the IASPEI was published in the Seismological Research Letters (Storchak et al., 2003). Various updates to the list were required due to progress in observational seismology and relevant changes in other observational standards. This article accommodates the advances made in the nomenclature since its last publication.

The new nomenclature partially modified and complemented an earlier one published in the old edition of the Manual of Seismological Observatory Practice (Willmore, 1979). It is more in tune with modern Earth and travel-time models. As opposed to former practice, the WG tried to make sure that the phase name generally reflects the type of wave and the path it has traveled. Accordingly, symbols for characterizing onset quality, polarity, etc., are no longer part of the phase name. The WG was also aware that seismic phases exist that are common in some regions but are only rarely or not found in other regions, such as Pb, PnPn, PbPb, etc.

The extended list of phase names as presented below reflects significantly increased detection capabilities of modern seismic sensors and sensor arrays, even of rather weak phases, which were rarely observed on the classical analog records. It also accounts for improved possibilities of proper phase identification by means of digital multi-channel data processing such as frequency-wave number (f-k) analysis and polarization filtering, by modeling the observations with synthetic seismograms or by showing on the records the theoretically predicted onset times of phases. Finally, the newly adopted IASPEI Seismic Format (ISF) (www.isc.ac.uk/doc/code/isf/isf.pdf) is much more flexible then the older formats previously used by the ISC, the NEIC, and other data centers. It also allows reporting, computer parsing, and archiving of phases with long or previously uncommon names. ISF also accepts complementary parameters such as onset quality, measured back azimuth and slowness, amplitudes and periods of other phases in addition to P and surface waves, for components other than vertical ones, and for instruments with nonstandard response characteristics.

This increased flexibility of the parameter-reporting format requires improved standardization, which limits the uncontrolled growth of incompatible and ambiguous parameter data. Therefore, the WG agreed on certain rules. They are outlined below prior to the listing of the standardized phase names. To facilitate the understanding of the phase names, ray diagrams are presented below. They have been calculated for local seismic sources on the basis of an average one-dimensional two-layer crustal model and for regional and teleseismic sources using the global 1D Earth model AK135 (Kennett et al., 1995).

Before assigning abbreviated shortcut seismic phase names, one should agree first on the language to be used and its rules. As in any other language, we need a suitable alphabet (here plain Latin letters), numbers (here Arabic numbers and +/− signs), an orthography, which regulates, for example, the use of capital and lower case letters, and a syntax, that describes the rules of correct order and mutual relationship of the language elements. One should be aware, however, that like any historically developed language, the seismological nomenclature will inevitably develop exceptions to the rules and depend on the context in which it is used. Although not fully documented below, some exceptions will be mentioned. Note that our efforts are mainly aimed at standardized names to be used in international data exchange so as to build up unique, unambiguous global databases for research. Many of the exceptions to the rules are related to specialized, mostly local research applications. The identification of related seismic phases often requires specialized procedures of data acquisition and processing that are not part of seismological routine data analysis. Also, many of these exceptional phases are rarely or never used in seismic event location, magnitude determination, source mechanism calculations, etc., which are the main tasks of international data centers. We focus, therefore, on phases that are particularly important for seismological data centers as well as for the refinement of regional and global Earth models on the basis of widely exchanged and accumulated parameter readings. In addition, we

added references to the first definition of some wave types and phase names.

Standard letters, signs, and syntax used for describing seismic phases

Capital letters

Individual capital letters that stand for primary types of seismic body waves include:

P:	Longitudinal wave that has traveled through the Earth's crust and mantle, from *undae primae* (Latin) = first waves (Borne, 1904)
K:	Longitudinal wave that has traveled through the Earth's outer core, K, from Kern (German) = core (Sohon, 1932; Bastings, 1934)
I:	Longitudinal wave that has traveled through the Earth's inner core (Jeffreys and Bullen, 1940)
S:	Transverse wave that has traveled through the Earth's crust and mantle, from *undae secundae* (Latin) = second waves (Borne, 1904)
T:	Wave that has partly traveled as sound wave in the sea, from *undae tertiae* (Latin) = third waves (Linehan, 1940)
J:	Transverse wave that has traveled through the Earth's inner core (Bullen, 1946)

Exceptions

- A capital letter N used in the nomenclature does not stand for a phase name but rather for the number of legs traveled (or N-1 reflections made) before reaching the station. N should usually follow the phase symbol to which it applies. For examples see syntax below.
- The lowercase letters p and s may stand, in the case of seismic events below the Earth's surface, for the relatively short *upgoing leg* of P or S waves, which continue, after reflection and possible conversion at the free surface, as downgoing P or S wave. Thus seismic depth phases (e.g., pP, sP, sS, pPP, sPP, pPKP, etc.) are uniquely defined. The identification and reporting of such phases is of utmost importance for source depth determination (Scrase, 1931; Stechschulte, 1932; Gutenberg et al., 1933; Macelwane et al., 1933).
- Many researchers working on detailed investigations of crustal and upper-mantle discontinuities denote both the up- and downgoing short legs of converted or multiply reflected P and S phases as lowercase letters p and s, respectively.

Individual or double capital letters that stand for surface waves include:

L:	(Relatively) long-period surface wave, unspecified, from *undae longae* (Latin) = long waves (Borne, 1904)
R:	Rayleigh waves (short- to very long-period waves in crust and upper mantle) (Angenheister, 1921)
Q:	Love waves, from Querwellen (German) = transverse waves (Angenheister, 1921)
G:	(Very long-period) global (mantle) Love waves, firstly observed and reported by Gutenberg and Richter (1934); in honor of Gutenberg, Byerly proposed the usage of G for these waves (Richter, 1958)
LR:	Long-period Rayleigh waves, usually relating to the Airy phase maximum in the surface wave train
LQ:	Long-period love waves

Lowercase letters and signs

Single lowercase letters generally specify the part of Earth's crust or upper mantle in which a phase has its turning point or at which discontinuity it has been reflected and/or eventually converted:

g:	Following the phase name characterizes waves "bottoming" (i.e., having their turning point in case of P or S body waves) or just travel (surface waves) within the upper ("granitic") Earth's crust (e.g., Pg, Sg; Rg), (Jeffreys, 1926)
b:	Following the phase name characterizes body waves bottoming in the lower ("basaltic") Earth's crust (Jeffreys, 1926) (e.g., Pb, Sb; alternative names for these phases are P*, S*, (Conrad, 1925))
n:	Following the phase name characterizes a P or S wave that is bottoming or traveling as head wave in the Earth's uppermost mantle (e.g., Pn, Sn), introduced after Andrija Mohorovičić discovered the Earth's crust and separated the crustal travel-time curve from the normal (=n) mantle phase (Mohorovičić, 1910)
m:	(Upward) reflections from the outer side of the Mohorovičić (Moho) discontinuity (e.g., PmP, SmS)
c:	Reflections from the outer side of the core-mantle boundary (CMB), usage proposed by James B. Macelwane (see Gutenberg, 1925)
i:	Reflections from the outer side of the inner core boundary (ICB)
z:	Reflections from a discontinuity (other than free surface, CMB or ICB) at depth z (measured in km). Upward reflections from the outer side of the discontinuity may additionally be complemented by a + sign (e.g., P410 + P; this, however, is not compulsory) while downward reflections from the inner side of the discontinuity must be complemented by a − sign (e.g., P660−P)

An exception from these rules is the use of lowercase p or s to indicate arrivals of longitudinal or transverse waves that were first radiated to go up toward the free surface to be reflected/converted back into the Earth as normal P or S waves (see near source surface reflections and conversions section of the phase list below).

Double lowercase letters following a capital letter phase name indicate the travel-time branch to which this phase belongs. Due to the geometry and velocity structure of the

Earth, the same type of seismic wave may develop a triplication of its travel-time curve with different, in some cases, well-separated, branches. Thus, it is customary to differentiate between different branches of core phases and their multiple reflections at the free surface or the CMB. Examples are PKPab, PKPbc, PKPdf, SKSac, SKKSac, etc. The separation of the different PKP branches with letters ab, bc, and df was introduced by Jeffreys and Bullen (1940).

Three lower case letters may follow a capital letter phase name to specify its character, e.g., as a forerunner (pre) to the main phase, caused by scattering (e.g., PKPpre) or as a diffracted wave extending the travel-time branch of the main phase into the outer core shadow (e.g., Pdif in the outer core shadow for P).

Syntax of generating complex phase names

Due to refraction, reflection, and conversion in the Earth, most phases have a complex path history before they reach the station. Accordingly, most phases cannot be described by a single capital letter code in a self-explanatory way. By combining the capital and lower case letters as mentioned above, one can describe the character of even rather complex refracted, reflected, or converted phases. The order of symbols (syntax) regulates the sequence of phase legs due to refraction, reflection, and conversion events in time (from left to right) and in space.

Examples for creating complex standard phase names

Traditional examples of complex phase names are as follows.

Refracted and converted refracted waves

- PKP is a pure refracted longitudinal wave. It has traveled the first part of its path as P through crust and mantle, the second through the outer core, and the third again as P through mantle and crust. An alternative name for PKP is P' (Angenheister, 1921), which should be read as "P prime."
- PKIKP (alternative to PKPdf) is also a pure refracted longitudinal wave. It has traveled the first part of its path as P through crust and mantle, the second through the outer core, the third through the inner core, and the fourth and fifth parts back again through outer core and mantle/crust.
- SKS is a converted refracted wave. It has traveled as a shear wave through crust and mantle, being converted into a longitudinal P wave when refracted into the outer core and converted back again into an S wave when entering the mantle.
- SKP or PKS are converted refracted waves with only one conversion from S to P when entering the core or from P to S when leaving the core, respectively.

Pure reflected waves

- In the case of (downward only) reflections at the free surface or from the inner side of the CMB, the phase symbol is just repeated, e.g., PP, SS (Geiger, 1909), PPP, SSS, KK, KKK, etc.
- In the case of (upward) reflections from the outer side of the Moho, the CMB, or the ICB, this is indicated by inserting symbols m, c, or i, respectively, between the phase symbols, e.g., PmP, PcP, ScS, PKiKP.
- Reflections from any other discontinuity in mantle or crust at depth z may be from the inner side (−; i.e., downward back into the mantle) or from the outer side (+; i.e., back toward the surface). To differentiate between these two possibilities, the sign has to follow z (or the respective number in km); for example, P410 + P or P660 − P.
- To abbreviate names of multi-leg phases due to repeated reflections, one can also write Phasename N. This type of abbreviation is customary in case of multiple phases with long phase names such as PmP2 for PmPPmP (free-surface reflection of PmP), SKS2 for SKSSKS (the alternative name for S'2, the free-surface reflection of SKS), PKP3 for PKPPKPPKP (double free-surface reflection of PKP; alternative name to P'3) or P4KP for PKKKKP (triple reflection of P at the inner side of the CMB).

Two additional notes are to be mentioned. First, PKP2 = PKPPKP are now alternative names for P'2 or P'P', respectively. This should not be mistaken for the old usage of PKP2 for PKPab. Secondly, in the case of multiple reflections from the inner side of the CMB, the WG followed the established tradition of placing the number N not after but in front of the related phase symbol K.

Reflected waves with conversion at the reflection point

In the case that a phase changes its character from P to S, or vice versa, one writes:

- PS (first leg P, second leg S) or SP (first leg P, second leg S) in the case of reflection/conversion from the free surface downward into the mantle (Geiger and Gutenberg, 1912a, 1912b).
- PmS or SmP, respectively, for reflection/conversion from the outer side of the Moho.
- PcS or ScP for reflection/conversion from the outer side of the CMB.
- Pz + S or Sz−P for reflection/conversion from the outer (+) side or inner (−) side, respectively, of a discontinuity at depth z. Note that the − is compulsory, the + is not.
- pS or sP reflection/conversion at the free surface for body waves with a direct upgoing first leg.

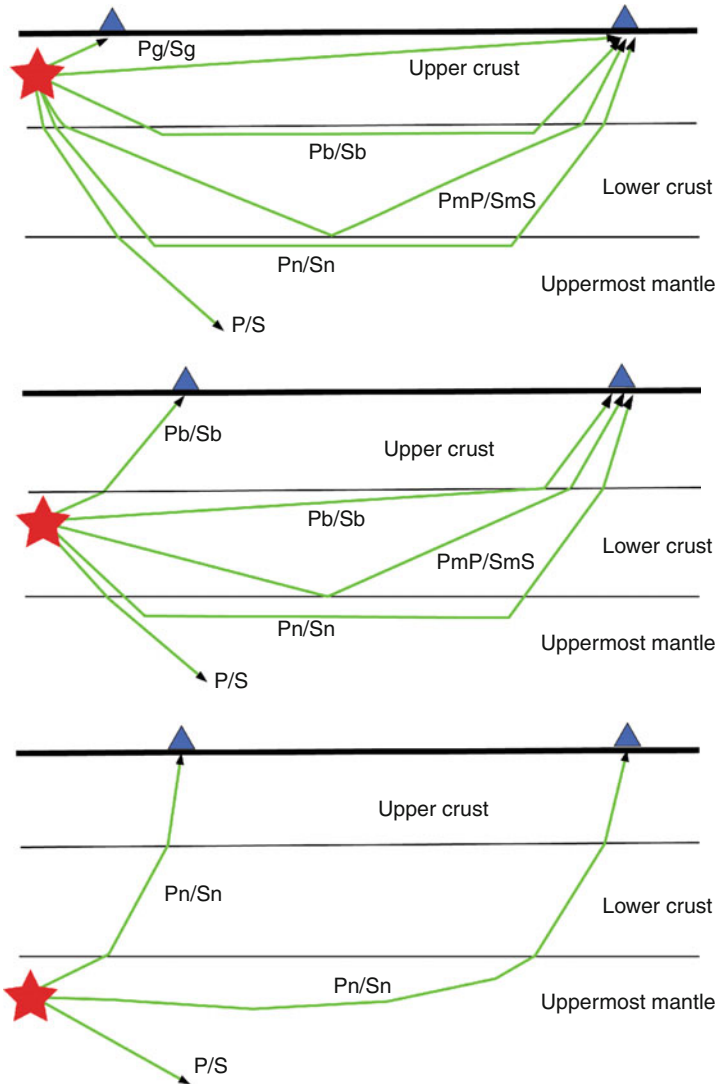
In this context, it is worth mentioning that mode conversion is impossible for reflections from the inner side of the CMB back into the outer core because the liquid outer core does not allow the propagation of S waves.

The WG determined the new IASPEI standard phase names along these lines and rules. Where these deviate from other traditionally used names, the latter are given

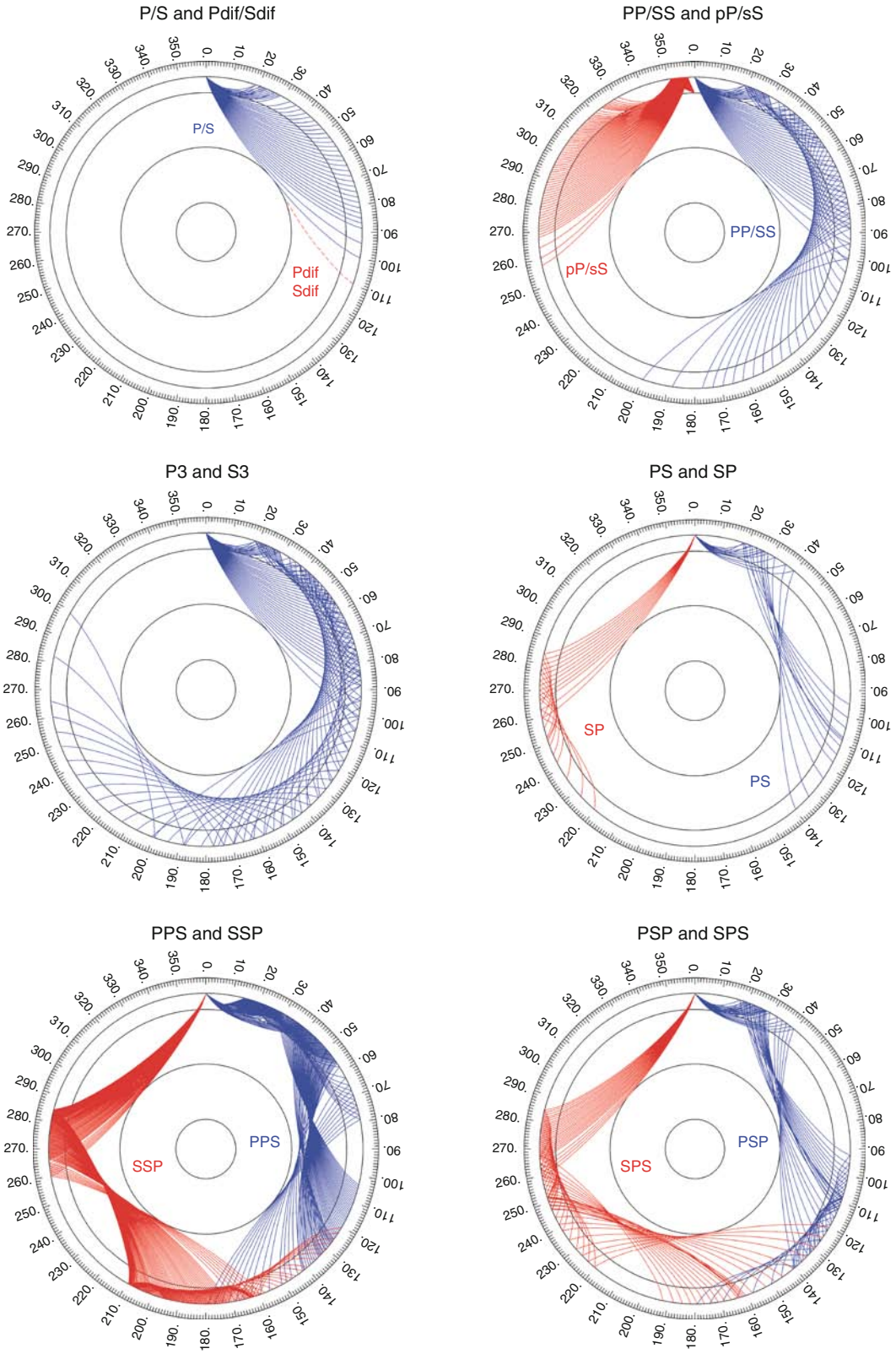
as well. Either the traditional names are still acceptable alternatives (alt) or they are old names (old), which should no longer be used.

Ray-path diagrams for some of the IASPEI standard phases

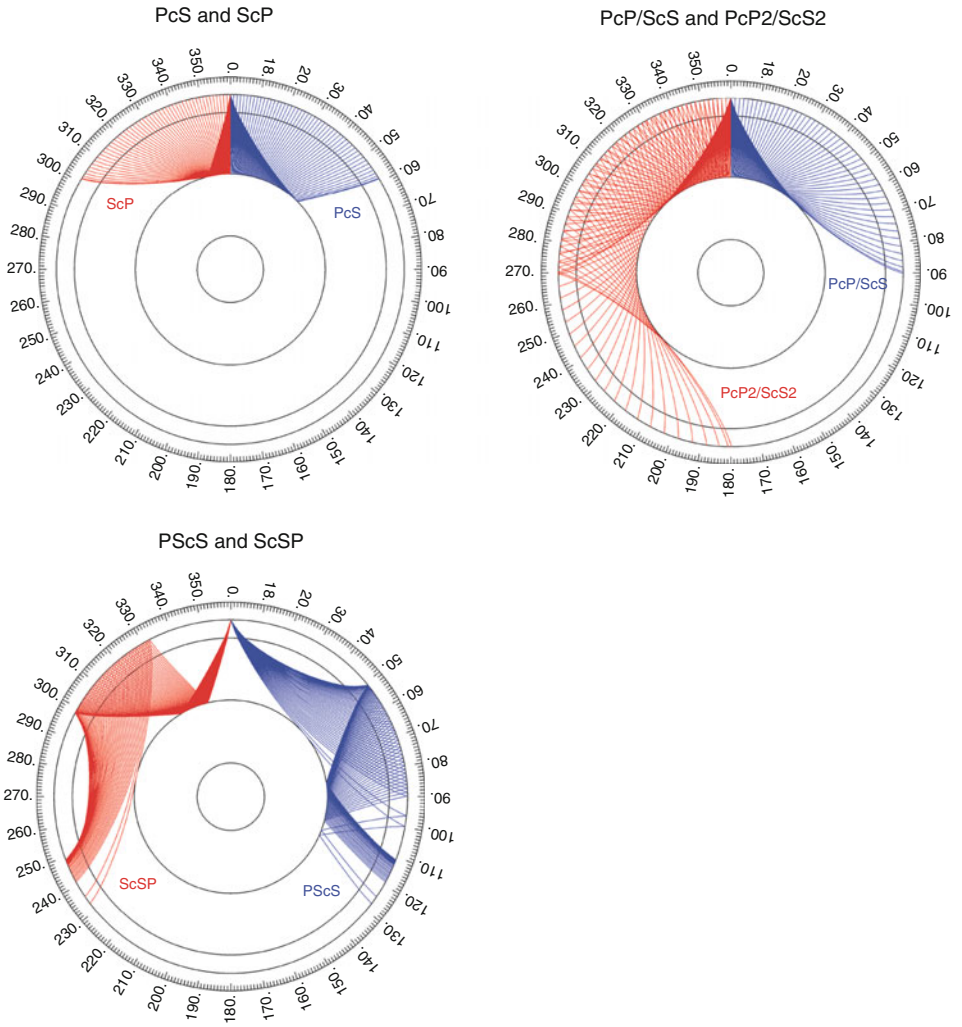
We show ray paths through the Earth for many of the mentioned phases. The three diagrams for crustal phases are sketches illustrating the principal ray paths in a two-layer crust (Figure 1). The rays in all other figures (Figures 2–6) were calculated by using the ray picture part of the WKBJ3 code (Chapman, 1978; Dey-Sarkar and Chapman, 1978); as the velocity model, we chose the standard Earth model AK135 (Kennett et al., 1995). For some types of P and S phases, the ray paths through the Earth are very similar because the velocity ratio V_P/V_S does not change enough to give very different ray pictures. In these cases, we calculated only the ray paths for the P-type ray (i.e., P, Pdif, pP, PP, P660P, P660−P, PcP, PcP2 and PcP4) and assume that the corresponding ray paths of the respective S-type phases are very similar. To show the different ray paths for phases with similar phase names, we show on many figures rays leaving the source once to the left and once to the right in different colors. The three most important discontinuities inside the Earth are indicated as black circles (i.e., the border between upper and lower mantle, the CMB, and the ICB).



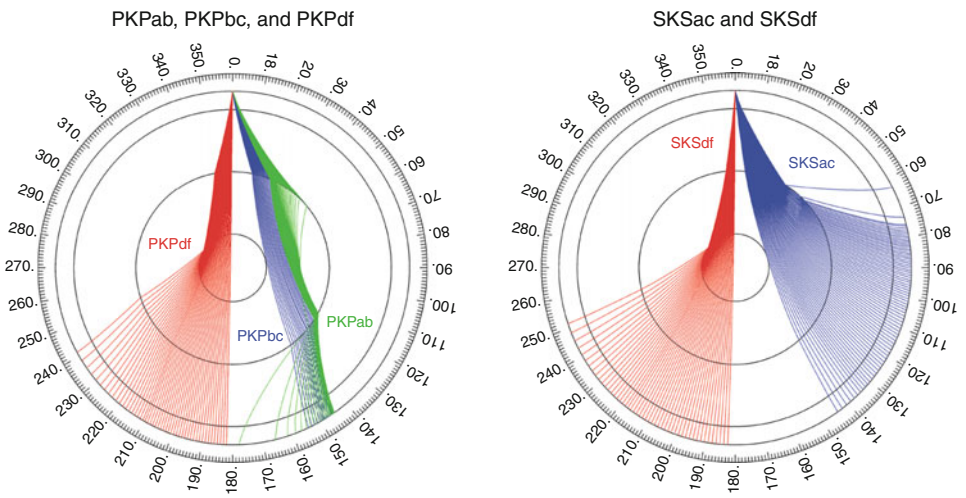
Seismic Phase Names: IASPEI Standard, Figure 1 Seismic "crustal phases" observed in the case of a two-layer crust in local and regional distance ranges ($0^\circ < D <$ approximately 20°) from the seismic source in the: (**a**) upper crust; (**b**) lower crust; and (**c**) uppermost mantle.



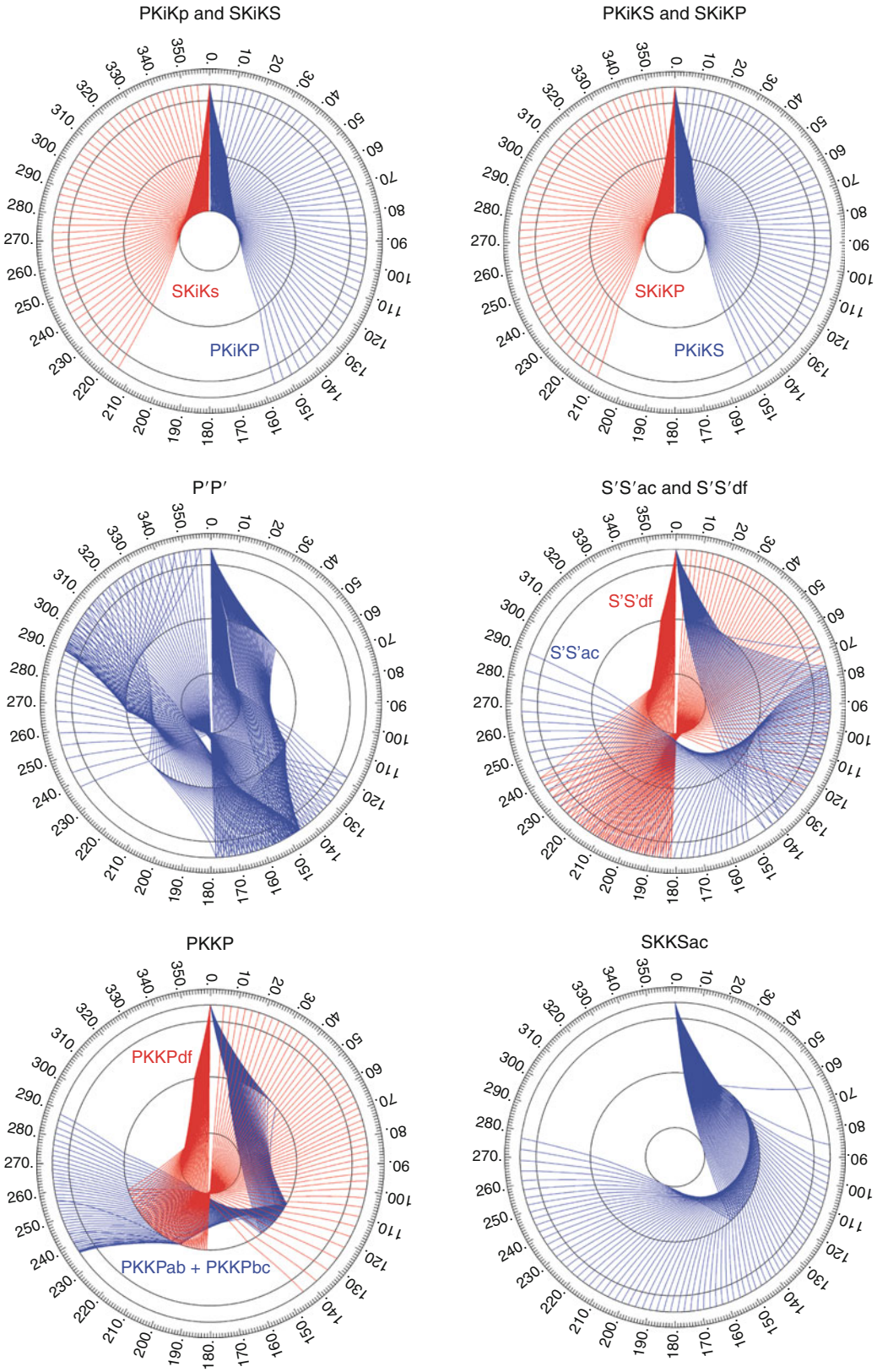
Seismic Phase Names: IASPEI Standard, Figure 2 Mantle phases observed at the teleseismic distances range (D > approximately 20°).



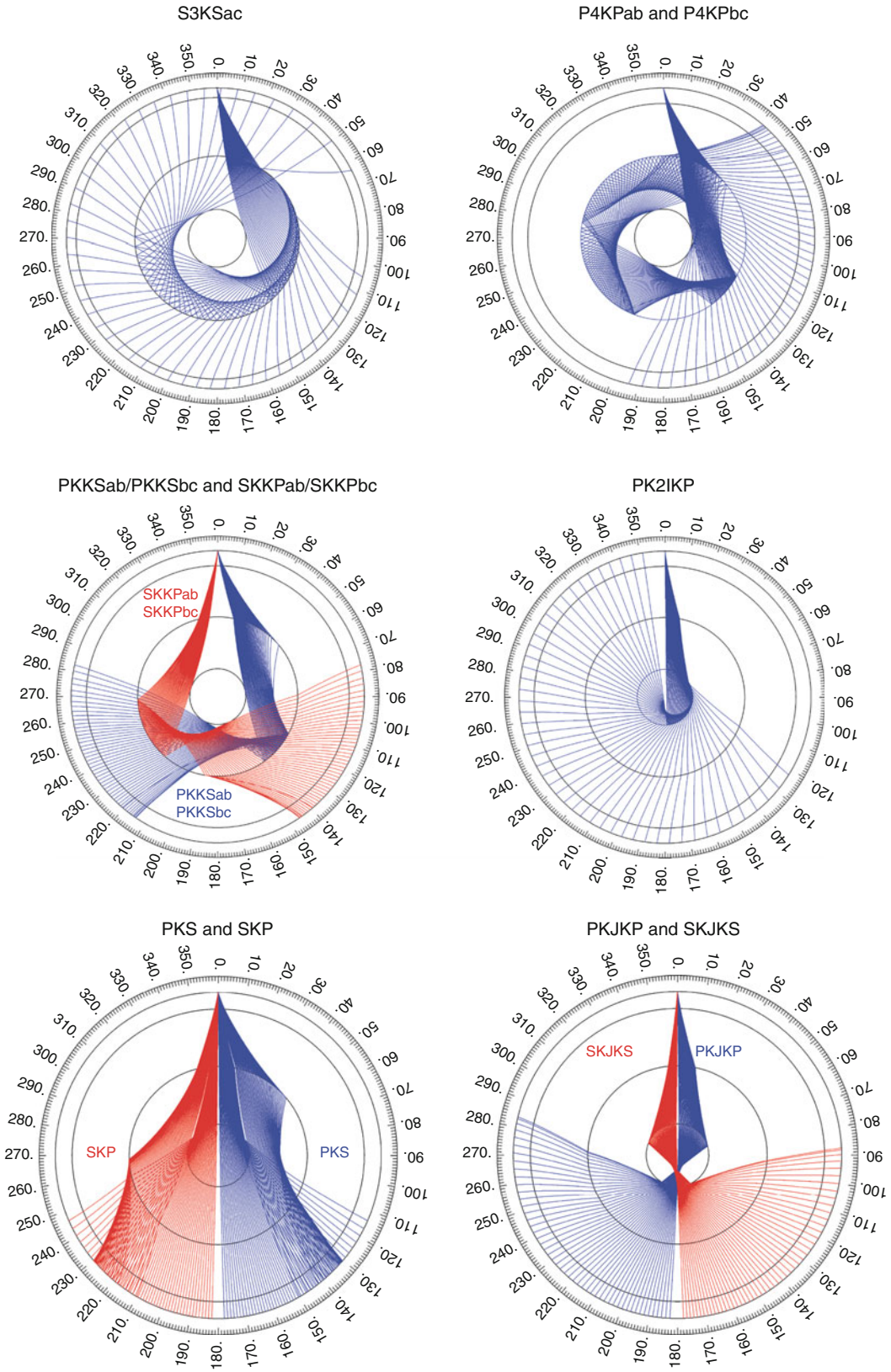
Seismic Phase Names: IASPEI Standard, Figure 3 Reflections from the Earth's core.



Seismic Phase Names: IASPEI Standard, Figure 4 Seismic rays of direct core phases.



Seismic Phase Names: IASPEI Standard, Figure 5 Seismic rays of single-reflected core phases.



Seismic Phase Names: IASPEI Standard, Figure 6 Seismic rays of multiple-reflected and converted core phases.

IASPEI standard seismic phase list

Crustal phases

Pg	At short distances, either an upgoing P wave from a source in the upper crust or a P wave bottoming in the upper crust. At larger distances also, arrivals caused by multiple P-wave reverberations inside the whole crust with a group velocity around 5.8 km/s
Pb	(alt:P*) Either an upgoing P wave from a source in the lower crust or a P wave bottoming in the lower crust
Pn	Any P wave bottoming in the uppermost mantle or an upgoing P wave from a source in the uppermost mantle
PnPn	Pn free-surface reflection
PgPg	Pg free-surface reflection
PmP	P reflection from the outer side of the Moho
PmPN	PmP multiple free-surface reflection; N is a positive integer. For example, PmP2 is PmPPmP
PmS	P to S reflection/conversion from the outer side of the Moho
Sg	At short distances, either an upgoing S wave from a source in the upper crust or an S wave bottoming in the upper crust. At larger distances also, arrivals caused by superposition of multiple S-wave reverberations and SV to P and/or P to SV conversions inside the whole crust
Sb	(alt:S*) Either an upgoing S wave from a source in the lower crust or an S wave bottoming in the lower crust
Sn	Any S wave bottoming in the uppermost mantle or an upgoing S wave from a source in the uppermost mantle
SnSn	Sn free-surface reflection
SgSg	Sg free-surface reflection
SmS	S reflection from the outer side of the Moho
SmSN	SmS multiple free-surface reflection; N is a positive integer. For example, SmS2 is SmSSmS
SmP	S to P reflection/conversion from the outer side of the Moho
Lg	A wave group observed at larger regional distances and caused by superposition of multiple S-wave reverberations and SV to P and/or P to SV conversions inside the whole crust. The maximum energy travels with a group velocity of approximately 3.5 km/s
Rg	Short-period crustal Rayleigh wave

Mantle phases

P	A longitudinal wave, bottoming below the uppermost mantle; also an upgoing longitudinal wave from a source below the uppermost mantle
PP	Free-surface reflection of a P wave leaving a source downward
PS	P, leaving a source downward, reflected as an S at the free surface. At shorter distances, the first leg is represented by a crustal P wave
PPP	analogous to PP
PPS	PP which is converted to S at the second reflection point on the free surface; travel time matches that of PSP
PSS	PS reflected at the free surface
PcP	P reflection from the core-mantle boundary (CMB)
PcS	P converted to S when reflected from the CMB
PcPN	PcP reflected from the free surface N − 1 times; N is a positive integer. For example PcP2 is PcPPcP
Pz + P	(alt:PzP) P reflection from outer side of a discontinuity at depth z; z may be a positive numerical value in km. For example, P660 + P is a P reflection from the top of the 660 km discontinuity
Pz−P	P reflection from inner side of a discontinuity at depth z. For example, P660 − P is a P reflection from below the 660 km discontinuity, which means it is precursory to PP
Pz + S	(alt:PzS) P converted to S when reflected from outer side of discontinuity at depth z
Pz−S	P converted to S when reflected from inner side of a discontinuity at depth z
PScS	P (leaving a source downward) to ScS reflection at the free surface
Pdif	(old:Pdiff) P diffracted along the CMB in the mantle
S	Shear wave, bottoming below the uppermost mantle; also an upgoing shear wave from a source below the uppermost mantle
SS	Free-surface reflection of an S wave leaving a source downward
SP	S, leaving a source downward, reflected as P at the free surface. At shorter distances, the second leg is represented by a crustal P wave
SSS	Analogous to SS
SSP	SS converted to P when reflected from the free surface; travel time matches that of SPS
SPP	SP reflected at the free surface
ScS	S reflection from the CMB
ScP	S converted to P when reflected from the CMB
ScSN	ScS multiple free-surface reflection; N is a positive integer. For example ScS2 is ScSScS
Sz + S	(alt:SzS) S reflection from outer side of a discontinuity at depth z; z may be a positive numerical value in km. For example S660 + S is an S reflection from the top of the 660 km discontinuity
Sz−S	S reflection from inner side of discontinuity at depth z. For example, S660 − S is an S reflection from below the 660 km discontinuity, which means it is precursory to SS
Sz + P	(alt:SzP) S converted to P when reflected from outer side of a discontinuity at depth z
Sz−P	S converted to P when reflected from inner side of a discontinuity at depth z
ScSP	ScS to P reflection at the free surface
Sdif	(old:Sdiff) S diffracted along the CMB in the mantle

Core phases

PKP	(alt:P') unspecified P wave bottoming in the core
PKPab	(old:PKP2) P wave bottoming in the upper outer core; ab indicates the retrograde branch of the PKP caustic
PKPbc	(old:PKP1) P wave bottoming in the lower outer core; bc indicates the prograde branch of the PKP caustic

PKPdf	(alt:PKIKP) P wave bottoming in the inner core
PKPpre	(old:PKhKP) a precursor to PKPdf due to scattering near or at the CMB
PKPdif	P wave diffracted at the inner core boundary (ICB) in the outer core
PKS	Unspecified P wave bottoming in the core and converting to S at the CMB
PKSab	PKS bottoming in the upper outer core
PKSbc	PKS bottoming in the lower outer core
PKSdf	PKS bottoming in the inner core
P'P'	(alt:PKPPKP) Free-surface reflection of PKP
P'N	(alt:PKPN) PKP reflected at the free surface N − 1 times; N is a positive integer. For example, P'3 is P'P'P'
P'z−P'	PKP reflected from inner side of a discontinuity at depth z outside the core, which means it is precursory to P'P'; z may be a positive numerical value in km
P'S'	(alt:PKPSKS) PKP converted to SKS when reflected from the free surface; other examples are P'PKS, P'SKP
PS'	(alt:PSKS) P (leaving a source downward) to SKS reflection at the free surface
PKKP	Unspecified P wave reflected once from the inner side of the CMB
PKKPab	PKKP bottoming in the upper outer core
PKKPbc	PKKP bottoming in the lower outer core
PKKPdf	PKKP bottoming in the inner core
PNKP	P wave reflected N − 1 times from inner side of the CMB; N is a positive integer
PKKPpre	A precursor to PKKPdf due to scattering near the CMB
PKiKP	P wave reflected from the inner core boundary (ICB)
PKNIKP	P wave reflected N − 1 times from the inner side of the ICB
PKJKP	P wave traversing the outer core as P and the inner core as S
PKKS	P wave reflected once from the inner side of the CMB and converted to S at the CMB
PKKSab	PKKS bottoming in the upper outer core
PKKSbc	PKKS bottoming in the lower outer core
PKKSdf	PKKS bottoming in the inner core
PcPP'	(alt:PcPPKP) PcP to PKP reflection at the free surface; other examples are PcPS', PcSP', PcSS', PcPSKP, PcSSKP
SKS	(alt:S') unspecified S wave traversing the core as P
SKSac	SKS bottoming in the outer core
SKSdf	(alt:SKIKS) SKS bottoming in the inner core
SPdifKS	(alt:SKPdifS) SKS wave with a segment of mantle-side Pdif at the source and/or the receiver side of the ray path
SKP	Unspecified S wave traversing the core and then the mantle as P
SKPab	SKP bottoming in the upper outer core
SKPbc	SKP bottoming in the lower outer core
SKPdf	SKP bottoming in the inner core
S'S'	(alt:SKSSKS) Free-surface reflection of SKS
S'N	SKS reflected at the free surface N − 1 times; N is a positive integer
S'z−S'	SKS reflected from inner side of discontinuity at depth z outside the core, which means it is precursory to S'S'; z may be a positive numerical value in km
S'P'	(alt:SKSPKP) SKS converted to PKP when reflected from the free surface; other examples are S'SKP, S'PKS
S'P	(alt:SKSP) SKS to P reflection at the free surface
SKKS	Unspecified S wave reflected once from inner side of the CMB
SKKSac	SKKS bottoming in the outer core
SKKSdf	SKKS bottoming in the inner core
SNKS	S wave reflected N − 1 times from inner side of the CMB; N is a positive integer
SKiKS	S wave traversing the outer core as P and reflected from the ICB
SKJKS	S wave traversing the outer core as P and the inner core as S
SKKP	S wave traversing the core as P with one reflection from the inner side of the CMB and then continuing as P in the mantle
SKKPab	SKKP bottoming in the upper outer core
SKKPbc	SKKP bottoming in the lower outer core
SKKPdf	SKKP bottoming in the inner core
ScSS'	(alt:ScSSKS) ScS to SKS reflection at the free surface; other examples are ScPS', ScSP', ScPP', ScSSKP, ScPSKP

Near source surface reflections and conversions (depth phases)

pPy	All P-type onsets (Py) as defined above, which resulted from reflection of an upgoing P wave at the free surface or an ocean bottom; WARNING: The character "y" is only a wild card for any seismic phase, which could be generated at the free surface. Examples are pP, pPKP, pPP, pPcP, etc.
sPy	All Py resulting from reflection of an upgoing S wave at the free surface or an ocean bottom; for example, sP, sPKP, sPP, sPcP, etc.
pSy	All S-type onsets (Sy) as defined above, which resulted from reflection of an upgoing P wave at the free surface or an ocean bottom. For example, pS, pSKS, pSS, pScP, etc.
sSy	All Sy resulting from reflection of an upgoing S wave at the free surface or an ocean bottom. For example, sSn, sSS, sScS, sSdif, etc.
pwPy	All Py resulting from reflection of an upgoing P wave at the ocean's free surface
pmPy	All Py resulting from reflection of an upgoing P wave from the inner side of the Moho

Surface waves

L	Unspecified long-period surface wave
LQ	Love wave
LR	Rayleigh wave
G	Mantle wave of Love type
GN	Mantle wave of Love type; N is integer and indicates wave packets traveling along the minor arcs (odd numbers) or major arc (even numbers) of the great circle
R	Mantle wave of Rayleigh type
RN	Mantle wave of Rayleigh type; N is integer and indicates wave packets traveling along the minor arcs (odd numbers) or major arc (even numbers) of the great circle

PL	Fundamental leaking mode following P onsets generated by coupling of P energy into the waveguide formed by the crust and upper mantle
SPL	S wave coupling into the PL waveguide; other examples are SSPL, SSSPL

Acoustic phases

H	A hydroacoustic wave from a source in the water, which couples in the ground
HPg	H phase converted to Pg at the receiver side
HSg	H phase converted to Sg at the receiver side
HRg	H phase converted to Rg at the receiver side
I	An atmospheric sound arrival, which couples in the ground
IPg	I phase converted to Pg at the receiver side
ISg	I phase converted to Sg at the receiver side
IRg	I phase converted to Rg at the receiver side
T	A tertiary wave. This is an acoustic wave from a source in the solid Earth, usually trapped in a low velocity oceanic water layer called the SOFAR channel (SOund Fixing And Ranging)
TPg	T phase converted to Pg at the receiver side
TSg	T phase converted to Sg at the receiver side
TRg	T phase converted to Rg at the receiver side

Amplitude measurements

The following set of amplitude measurement names refers to the IASPEI Magnitude Standard (see www.iaspei.org/commissions/CSOI/Summary_of_WG_recommendations.pdf), compliance to which is indicated by the presence of leading letter I. The absence of leading letter I indicates that a measurement is nonstandard. Letter A indicates a measurement in nm made on a displacement seismogram, whereas letter V indicates a measurement in nm/s made on a velocity seismogram.

IAML	Displacement amplitude measured according to the IASPEI standard for local magnitude M_L
IAMs_20	Displacement amplitude measured according to IASPEI standard for surface-wave magnitude $M_S(20)$
IVMs_BB	Velocity amplitude measured according to IASPEI standard for broadband surface-wave magnitude $M_S(BB)$
IAmb	Displacement amplitude measured according to IASPEI standard for short-period teleseismic body-wave magnitude m_b
IVmB_BB	Velocity amplitude measured according to IASPEI standard for broadband teleseismic body-wave magnitude $m_B(BB)$
AX_IN	Displacement amplitude of phase of type X (e.g., PP, S, etc.), measured on an instrument of type IN (e.g., SP, short-period; LP, long-period; BB, broadband)
VX_IN	Velocity amplitude of phase of type X and instrument of type IN (as above)
A	Unspecified displacement amplitude measurement
V	Unspecified velocity amplitude measurement
AML	Displacement amplitude measurement for nonstandard local magnitude
AMs	Displacement amplitude measurement for nonstandard surface-wave magnitude
Amb	Displacement amplitude measurement for nonstandard short-period body-wave magnitude
AmB	Displacement amplitude measurement for nonstandard medium to long-period body-wave magnitude
END	Time of visible end of record for duration magnitude

Unidentified arrivals

x	(old: i, e, NULL) unidentified arrival
rx	(old: i, e, NULL) unidentified regional arrival
tx	(old: i, e, NULL) unidentified teleseismic arrival
Px	(old: i, e, NULL, (P), P?) unidentified arrival of P-type
Sx	(old: i, e, NULL, (S), S?) unidentified arrival of S-type

Bibliography

Angenheister, G. H., 1921. Beobachtungen an pazifischen Beben, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse,* 113–146.

Bastings, L., 1934. Shear waves through the Earth's core. *Nature*, **134**, 216–217.

Borne, G. von dem, 1904. Seismische Registrierungen in Göttingen, Juli bis Dezember 1903, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse,* 440–464.

Bullen, K. E., 1946. A hypothesis on compressibility at pressures of the order of a million atmospheres. *Nature*, **157**, 405.

Chapman, C. H., 1978. A new method for computing synthetic seismograms. *Geophysical Journal of the Royal Astronomical Society*, **54**, 481–518.

Conrad, V., 1925. Laufzeitkurven des Tauernbebens vom 28. November, 1923, *Mitteilungen der Erdbeben-Kommission der Akademie der Wissenschaften in Wien*, Neue Folge. **59**, 23.

Dey-Sarkar, S. K., and Chapman, C. H., 1978. A simple method for the computation of body wave seismograms. *Bulletin of the Seismological Society of America*, **68**, 1577–1593.

Geiger, L., 1909. Seismische Registrierungen in Göttingen im Jahre 1907 mit einem Vorwort über die Bearbeitung der Erdbebendiagramme, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse*, 107–151.

Geiger, L., and Gutenberg, B., 1912a. Konstitution des Erdinnern, erschlossen aus der Intensität longitudinaler und transversaler Erdbebenwellen, und einige Beobachtungen an den Vorläufern. *Physikalische Zeitschrift*, **13**, 115–118.

Geiger, L., and Gutenberg, B., 1912b. Ueber Erdbebenwellen. VI. Konstitution des Erdinnern, erschlossen aus der Intensität longitudinaler und transversaler Erdbebenwellen, und einige Beobachtungen an den Vorläufern, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse*, 623–675.

Gutenberg, B., 1925. Bearbeitung von Aufzeichnungen einiger Weltbeben. *Abhandlungen der Senckenbergischen Naturforschenden Gesellschaft*, **40**, 57–88.

Gutenberg, B., and Richter, C. F., 1934. On seismic waves (first paper). *Gerlands Beitrage zur Geophysik*, **43**, 56–133.
Gutenberg, B., Wood, H. O., and Richter, C. F., 1933. Re suggestion by Dr. Harold Jeffreys regarding $\bar{P}$ and Pg. *Gerlands Beitrage zur Geophysik*, **40**, 97–98.
Jeffreys, H., 1926. On near earthquakes. *Monthly Notices of the Royal Astronomical Society. Geophysical Supplement*, **1**, 385–402.
Jeffreys, H., and Bullen, K. E., 1940. *Seismological Tables*. London: British Association for the Advancement of Science, Gray Milne Trust, p. 50.
Kennett, B. L. N., Engdahl, E. R., and Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, **122**, 108–124.
Linehan, D., 1940. Earthquakes in the West Indian region. *Transactions of the American Geophysical Union*, **30**, 229–232.
Macelwane J. B., Brunner, G. J., and Joliat, J. S., 1933. Re suggestion by Doctor Harold Jeffreys and others regarding $\bar{P}$ and Pg, *Gerlands Beiträge zur Geophysik*, **40**, 98.
Mohorovičić, A., 1910. Potres od 8. X 1909, *God. Izvjeste Zagr. met. Ops. Zag. 1909*, Zagreb. (Das Beben vom 8. X 1909, *Jahrbuch des meteorologischen Observatoriums in Zagreb für das Jahr 1909*, **9**, 4, 1–63).
Richter, C. F., 1958. *Elementary Seismology*. San Francisco and London: W. H. Freeman, p. 768.
Scrase, F. J., 1931. The reflected waves from deep focus earthquakes. *Proceedings of the Royal Society of London. Series A*, **A-132**, 213–235.
Sohon, F. W., 1932. *Seismometry*, Part II of Macelwane, J. B., and Sohon, F. W., 1932. *Introduction to Theoretical Seismology*, New York, 149 pp.
Stechschulte, V. C., 1932. The Japanese earthquake of March 29, 1928. *Bulletin of the Seismological Society of America*, **22**, 81–137.
Storchak, D. A., Bormann, P., and Schweitzer, J., 2002. Standard nomenclature of seismic phases. In: Bormann, P. (eds.), *New Manual of Seismological Observatory Practice,* GeoForschungsZentrum, Potsdam, Vol. 2, IS2.1, 1–18.
Storchak, D. A., Schweitzer, J., and Bormann, P., 2003. The IASPEI standard seismic phase list. *Seismological Research Letters*, **74**, 761–772.
Willmore, P. L., 1979. *Manual of Seismological Observatory Practice*, World Data Center A for Solid Earth Geophysics, Report SE-20, September 1979, Boulder, Colorado, 165 pp.

Cross-references

Seismogram Interpretation

SEISMIC PROPERTIES OF ROCKS

Nikolas I. Christensen
Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC, Canada

Synonyms

Rock P and S velocities

Definition

Compressional (*P*) *waves.* Seismic waves in which the rock particles vibrate parallel to the direction of wave propagation.

Shear (*S*) *waves.* Seismic waves in which the rock particles vibrate perpendicular to the direction of wave propagation.

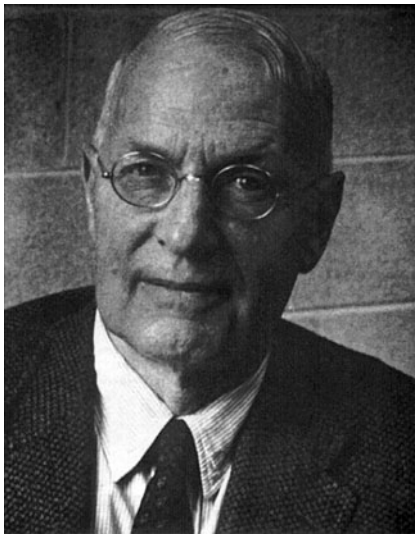
Poisson's ratio (σ). The ratio of the lateral unit strain to the longitudinal unit strain in a body that has been stressed longitudinally within its elastic limit ($2\sigma = (R^2 - 2)/(R^2 - 1)$ where $R = Vp/Vs$).

Transversely isotropic. Anisotropic solids with a single symmetry axis. Rock symmetry axes are usually normal to foliation, cleavage, or layering.

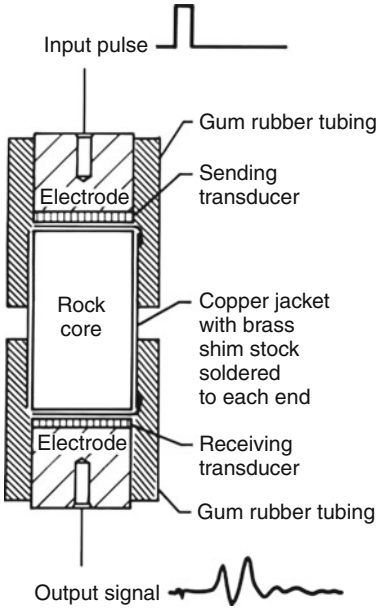
Introduction

Although many disciplines have contributed significantly to our knowledge of the Earth's interior, none has a resolution comparable to seismology. For nearly 6 decades seismic studies have provided geophysicists with worldwide information on crustal and upper mantle compressional (P) and shear (S) wave velocities. Significant data have recently become available on velocity gradients, velocity reversals, compressional and shear wave velocity ratios, and anisotropy in the form of azimuthal variations of compressional wave velocities, as well as shear wave splitting. Reflections within the crust and mantle originate from contrasts of acoustic impedances, defined as products of velocity and density. The interpretation of this seismic data requires detailed knowledge of rock velocities provided by laboratory techniques to a precision at least comparable with that of seismic measurements. In particular, to infer composition of the Earth's upper 30–50 km, the "crust," requires studies of the elasticity of rocks at conditions approaching those that exist at these depths. Of fundamental importance is the presence of mean compressive stress and temperature increasing with depth and on the average reaching about 1 GPa and 500 °C at the base of the crust. Because of this, the most relevant velocity measurements for identifying probable rock types within the crust have been measurements at elevated pressures and temperatures. These measurements often allow the seismologist to infer mineralogy, porosity, the nature of fluids occupying pore spaces, temperature at depth, and present or paleolithospheric stress based on mineral and crack orientations.

Francis Birch (Figure 1) was the pioneer in the study of rock velocities. In addition to his laboratory work on physical properties of rocks and minerals at high pressures and temperatures, he was well known for his studies of heat flow and theoretical work on the composition of the Earth's interior. Two of his benchmark papers on compressional wave velocities in rocks (Birch, 1960, 1961) set the stage for modern experimental studies of rock elasticity and have been frequently cited during the past 5 decades. These papers for the first time provided information on compressional wave velocities for many common rock types, as well as major findings on their anisotropies and relations to density. It is interesting to note that these measurements were carried out to pressures of 1 GPa, a pressure at which even today only a limited number of



Seismic Properties of Rocks, Figure 1 Francis Birch (1903–1992), a pioneer in rock physics research.



Seismic Properties of Rocks, Figure 2 Transducer and rock core assembly for velocity measurements at elevated pressures.

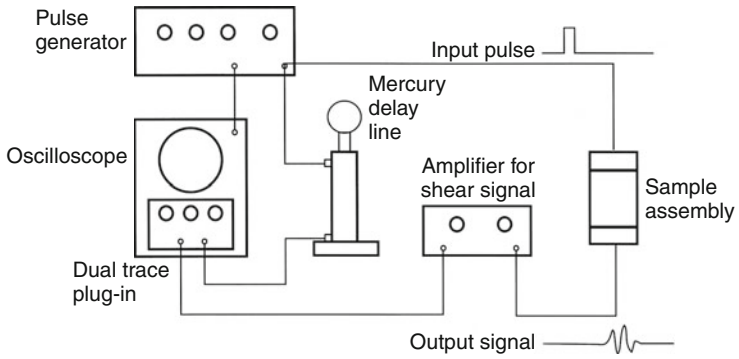
laboratories have been able to generate for modern rock seismic velocity measurements.

Measurement techniques

Rock velocities are usually measured in the laboratory using the pulse transmission technique. The transit time of either a compressional or shear wave is measured along the axis of a cylindrical rock specimen of known length. The cores are usually taken from rock samples using a 2.54 cm inner diameter diamond coring bit. The cores are trimmed and ground flat and parallel on a diamond grinding disk. The volume of each core is obtained from the length and diameter. The cores are weighed and densities are calculated from their masses and dimensions. The cores are then fitted with a copper jacket to prevent penetration of high-pressure oil into the rock samples. For measurements at high temperatures, where gas is the pressure medium, the samples are usually encased in stainless steel. Transducers are placed on the ends of the rock core (Figure 2). Compressional and shear waves are often generated by means of lead zirconate titanate (PZT) and AC cut quartz transducers with resonant frequencies of 1 MHz. The sending transducer converts the input, an electrical pulse of 50–500 V and 0.1–10 μs width, to a mechanical signal, which is transmitted through the rock. The receiving transducer changes the wave to an electrical pulse, which is amplified and displayed on an oscilloscope screen (Figure 3). Once the system is calibrated for time delays, the travel time through the specimen is determined directly by a computer or with the use of a mercury delay line. The major advantage of the delay line is that it increases the precision, especially for signals with slow rise times, because the gradual onset of the first arrival from the sample is approximated by the delay line. The velocity is the ratio of the length of the specimen to the travel time of the compressional or shear wave. The total error limits for Vp and Vs are estimated to be less than 1%. Interfacing the pressure system with a computer for data acquisition and storage permits automatic calculations of velocities as successive readings are taken. Using a least-squares routine, the computer fits a curve to the data points and calculates velocities for selected pressures. A velocity versus pressure curve is plotted along with recorded data points. Sample length, density, measured pressure velocity pairs, traces of the waveforms at selected pressures, the curve fit equations, and calculated pressure velocity pairs are recorded and stored digitally.

Hydrostatic pressure generating systems capable of producing true hydrostatic pressures as high as 3 GPa, equivalent to a depth of approximately 100 km, have been used for rock velocity measurements. Low viscosity synthetic petroleum and argon for high temperature measurements are frequently used as pressure media. An alternate technique for obtaining velocities under quasi-hydrostatic conditions has used a triaxial press with cubic samples. Transducers are placed on the six pistons and corrections are made for travel times through the pistons. Rock velocities obtained using this technique have provided valuable information on the effect of temperature on velocity, but have been limited to pressures of 0.6 GPa.

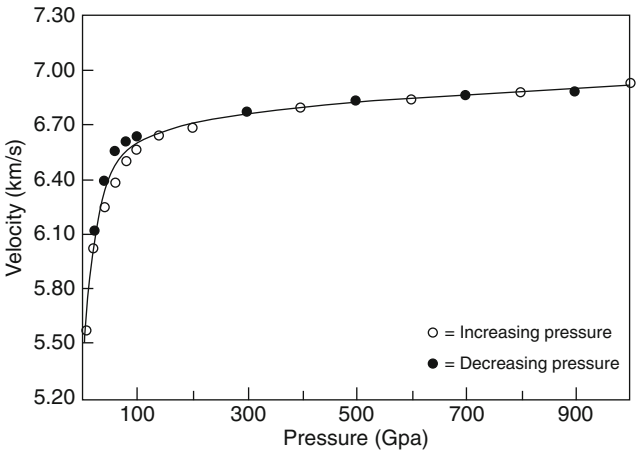
The behavior of a rocks velocity as a function of pressure is primarily dependent upon mineralogy and porosity. Many igneous and metamorphic rocks have porosities of the order of a few tenths of 1%, which are present as thin openings between grain boundaries. As pressure is applied to the rock, the cracks close and velocities increase. Once the cracks close any increase in velocity with increasing pressure is related to the intrinsic effects of pressure on



Seismic Properties of Rocks, Figure 3 Electronics for velocity measurements using a mercury delay line.

the mineral velocities. This is illustrated in Figure 4 for a garnet granulite. Velocities first increase rapidly over the first 100 MPa as cracks close and then increase slowly as pressure is increased. Also the velocity determined at a given pressure depends upon whether the pressure is approached from lower or higher pressure (Figure 4). This hysteresis is usually quite small if sufficient time is taken for measurements between pressure increments.

A considerable number of investigations have also focused on the influence of temperature on rock velocities. These studies have used either resonance techniques or more frequently the pulse transmission method. Early studies demonstrated that the application of temperature to rock at atmospheric pressure results in the creation of cracks that often permanently damage the rock and dramatically lower velocities. Thus reliable measurements of the temperature derivatives of velocities are obtained only at confining pressures high enough to prevent crack formation. At elevated confining pressures $\delta Vp/\delta T$ for common rocks often ranges from -0.3×10^{-3} to -0.6×10^{-3} km/s/°C and $\delta Vs/\delta T$ varies between -0.2×10^{-3} and -0.4×10^{-3} km/s/°C.



Seismic Properties of Rocks, Figure 4 Compressional wave velocity measurements as a function of confining pressure for a mafic granulite.

Rock velocities

Seismic velocities have been measured for practically all igneous and metamorphic rock types believed to be important constituents of the lithosphere. Because rock classification schemes allow for considerable variations in mineralogy for a given rock type, many rocks have wide ranges in elastic properties. However, some lithologies, such as the monomineralic rocks hornblendite and dunite with little or no alteration have fairly well-defined velocities. For detailed summaries of rock seismic properties the reader is referred to the compilations of Birch (1960), Christensen (1982, 1996), Gerbande (1982), Holbrook et al. (1992), Rudnick and Fountain (1995), and Mavko et al. (1998).

Table 1 contains average velocities, in the order of increasing compressional wave velocity, for several common igneous and metamorphic rocks. Volcanic rocks usually have lower velocities than their plutonic equivalents. This is due to the presence of glass, abundant alteration products and vesicles in volcanic rocks, all of which have lower velocities. In general, for a given composition, velocity increases with increasing metamorphic grade. For example, mica and quartz bearing schists have higher velocities than slates and phyllites. Low-grade metamorphosed basalts have lower velocities than higher grade amphibolite and mafic granulite. Eclogites have the highest velocity of mafic rocks. Note that shear velocities are relatively high in quartzites and low in serpentinites.

Early attempts to infer crustal composition by comparing laboratory and field derived velocities relied primarily on compressional wave velocities. However correlations between compressional wave velocity and composition are limited due to the similar velocities of many common crustal rock types. Because of this nonuniqueness of compressional wave velocity laboratory and field data comparisons, many recent studies have focused on investigations of crustal composition using both compressional and shear

wave velocities. In these studies the ratio Vp/Vs or Poisson's ratio (σ) calculated from Vp/Vs have resolved some of the ambiguities.

The values of Vp/Vs and σ, assuming isotropic elasticity, are given in Table 1 for several common igneous and metamorphic rocks at 1 GPa. This high-pressure eliminates cracks so the values only reflect mineralogy. The relatively low σ for quartzites (0.10) agrees well with isotropic aggregate calculations based on the elastic constants of single crystal quartz. Anorthosites, on the other hand, have relatively high Poisson's ratios (~0.31). As expected, values for the granites and granitic gneisses, consisting primarily of quartz and feldspar, fall in between those of quartzites and anorthosites and are relatively low. Thus, crustal regions where field measured values of $\sigma \leq 0.25$ are observed, are likely quartz-rich. Serpentinites containing lizardite, the variety of serpentine stable at crustal PT conditions, have extremely high values of Poisson's ratio (0.36), whereas unaltered dunites and peridotites have Poisson's ratios in the range of 0.25–0.26. Partially serpentinized dunites and peridotites have Poisson's ratios that fall between these limiting values. Laboratory measurements have established a well-defined relationship between Poisson's ratio, percent serpentinization and density. Changes in Poisson's ratio with progressive metamorphism of mafic igneous and pelitic rocks are considerably more complicated than the above examples (Christensen, 1996).

Seismic Properties of Rocks, Table 1 Average compressional (Vp) and shear (Vs) wave velocities, velocity ratios (Vp/Vs), and Poisson's ratios (σ) at 1 GPa for common rock types Christensen (1996)

Rock	Vp (km/s)	Vs (km/s)	Vp/Vs	σ
Serpentinite	5.607	2.606	2.152	0.36
Andesite	5.940	3.177	1.870	0.30
Quartzite	6.091	4.054	1.502	0.10
Basalt	6.118	3.291	1.859	0.30
Granitic gneiss	6.271	3.627	1.729	0.25
Granite-Granodiorite	6.372	3.726	1.710	0.24
Tonalite gneiss	6.366	3.636	1.751	0.26
Slate	6.379	3.432	1.858	0.30
Phyllite	6.398	3.608	1.774	0.27
Mica quartz schist	6.523	3.654	1.785	0.27
Zeolite facies basalt	6.530	3.493	1.869	0.30
Diorite	6.675	3.756	1.777	0.27
Diabase	6.814	3.766	1.809	0.28
Greenschist facies basalt	6.983	3.955	1.766	0.26
Marble	6.985	3.794	1.841	0.29
Mafic granulite	7.000	3.849	1.818	0.28
Amphibolite	7.046	3.987	1.767	0.26
Anorthosite	7.124	3.717	1.917	0.31
Gabbro	7.299	3.929	1.858	0.30
Pyroxenite	7.935	4.519	1.756	0.26
Eclogite	8.198	4.594	1.785	0.27
Dunite	8.399	4.783	1.756	0.26

Velocity anisotropy

Most crustal and upper mantle rocks show some degree of velocity anisotropy, which can originate from several processes. Laminar flow within magma and lava will orient elongate crystals such as feldspars along flow directions. Tabular sediments may settle preferentially and anisotropy may be enhanced by sediment compaction. Plastic flow and recrystallization during metamorphism often produce strong mineral orientations parallel to foliation and banding. In shallow crustal rocks oriented cracks producing anisotropy often originate from differential principle stresses. Anisotropy observed in laboratory measurements at pressures above approximately 100 MPa, where cracks are closed, often originate from preferred orientations of highly anisotropic minerals such as micas, amphiboles, pyroxenes, and olivine.

In general, for a given propagation direction in an anisotropic rock there are three waves, one compressional and two shear. Their vibration directions form an orthogonal set, which usually are not parallel or perpendicular to their propagation direction. Compressional and shear wave velocities vary with propagation direction, and two shear waves travel in a given direction through the rock with different velocities. This latter property of anisotropic rocks, termed shear wave splitting, was first recognized in laboratory studies and has been observed by field studies in several crustal and upper mantle regions.

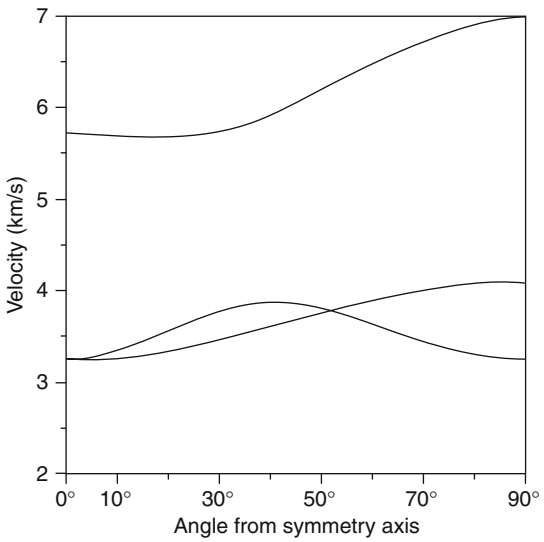
Anisotropy is usually studied in the laboratory by taking cores from a rock in different directions. Following the early investigations of Birch (1960, 1961) it is common practice to take three mutually perpendicular cores for velocity measurements. If the rock has a planar structure, such as a foliation, cleavage, or bedding, two cores are oriented with their axes within the planar structure. One of these cores is oriented parallel to a lineation, if present. In general, three velocities are measured per core: the compressional wave velocity, the velocity of the shear wave vibrating parallel to the layering, and the velocity of the shear wave vibrating in a plane perpendicular to layering. For cores taken perpendicular to layering, shear wave velocities are measured with vibration directions parallel to the axes of the cores taken in the layering.

It has been demonstrated that the above procedure provides information on maximum compressional wave anisotropy and maximum shear wave splitting. In general, the highest compressional wave velocities propagate in the plane of the foliation and parallel to lineations. Maximum shear wave splitting is observed for propagation within the planar structures. At near- normal incidence there is often minimal shear wave splitting.

Velocity measurements using multiple cores have provided detailed information on anisotropic velocities for non-axial propagation (e.g., Johnston and Christensen, 1995; Christensen and Okaya, 2007). The number of measurements necessary to completely describe wave propagation depends on the symmetry of the rock fabric. In the most general anisotropic elastic solid, 21 independent

constants are required to describe the equations of motion. Examples of materials with this type of behavior are minerals possessing triclinic symmetry such as kyanite and some feldspars. For many rocks, the existence of symmetry elements in the elastic properties leads to the vanishing of some elastic constants along with simple algebraic relations between others (e.g., Auld, 1990). Some crustal and upper mantle metamorphic rocks behave as elastic solids with orthorhombic symmetry, which require nine independent constants to describe the elastic tensors. Shales and many low to medium grade metamorphic rocks often have well-developed bedding or foliation and behave as transversely isotropic elastic solids (hexagonal symmetry with the symmetry axis normal to bedding or foliation). Transversely isotropic solids have five independent elastic constants, which can be calculated from five independent velocity measurements (two compressional wave velocities, one quasi-compressional wave velocity, and two shear wave velocities) and density. In an isotropic solid, only two independent constants are required for a complete description of elastic behavior and wave velocities are independent of propagation direction.

To describe three-dimensional wave propagation in anisotropic rocks, phase velocity surfaces can be calculated using the Kelvin-Christoffel equations (e.g., Auld, 1990) and elastic constants can be determined from velocity and density measurements. For transversely isotropic rocks, these surfaces describe variations in velocity as a function of angle to the bedding or foliation (Figure 5). Three velocity surfaces are calculated, one for the quasi-compressional wave, one for the shear wave vibrating parallel to the planar structure, and one for the quasi-shear wave vibrating in a plane perpendicular to the foliation or bedding. For propagation parallel and perpendicular to the foliation all wave modes are pure. The velocities shown in Figure 5 were calculated from velocity measurements at 600 MPa for a transversely isotropic quartz -mica schist from South Island, New Zealand. They show several important features about elastic wave propagation in this rock, which are typical of many foliated rocks. First, compressional wave velocities do not increase significantly until propagation directions greater than about 45° from foliation normal are reached. At larger angles, compressional wave velocity increases rapidly and reaches a maximum for propagation parallel to the foliation. Shear wave singularities (directions in which two shear waves have equal velocities) occur for propagation parallel to and at approximately 50° to the symmetry axis. Shear wave splitting occurs for all other propagation directions and reaches to maximum of 90° from the normal to the foliation.



Seismic Properties of Rocks, Figure 5 Compressional wave anisotropy (*upper curve*) and shear wave splitting (*lower curves*) for a transversely isotropic quartz-mica schist.

Summary

Beginning with the compressional wave velocity measurements of Birch (1960, 1961), much has been learned about elastic wave propagation in rocks. Compressional wave velocities are now available for most common rock types at pressures existing in the continental crust and uppermost mantle. Additional important contributions include laboratory measurements of shear wave velocities, velocity ratios, and the influence of temperature on velocities. Estimates of crustal composition from compressional wave velocities are nonunique, but this ambiguity is often reduced by complimentary shear wave velocity (and Poisson's ratio) observations. Recent field studies have found that seismic anisotropy is an important crustal feature. Thus systematic laboratory studies of compressional wave anisotropy and shear wave splitting will be critical in understanding crustal composition and deformation, just as they have been in investigations of the upper mantle.

Bibliography

Auld, B. A., 1990. *Acoustic Fields and Waves in Solids*. Malabar: Robert E. Krieger, Vol. 1.

Birch, F., 1960. The velocity of compressional waves in rocks to 10 kilobars, 1. *Journal of Geophysical Research*, **65**, 1083–1102.

Birch, F., 1961. The velocity of compressional waves in rocks to 10 kilobars, 2. *Journal of Geophysical Research*, **66**, 2199–2224.

Christensen, N. I., 1982. Seismic velocities. In Carmichael, R. S. (ed.), *Handbook of Physical Properties of Rocks*, 2nd edn. Boca Raton: CRC, pp. 1–228.

Christensen, N. I., 1996. Poisson's ratio and crustal seismology. *Journal of Geophysical Research*, **101**, 3139–3156.

Christensen, N. I., and Okaya, D. A., 2007. Compressional and shear wave velocities in South Island, New Zealand rocks and their application to the interpretation of seismological models of the New Zealand crust. In Okaya, D., Stern, T., and Davey, F. (eds.), *A Continental Plate Boundary: Tectonics at South Island, New Zealand*. Washington, DC: American Geophysical Union. American Geophysical Union geophysical monograph 175, pp. 125–155.

Gerbande, H., 1982. Elastic wave velocities and constants of elasticity of rocks and rock forming minerals. In Angenheister, G. (ed.), *Physical Properties of Rocks*. Springer: Landolt-Bornstein, pp. 1–140.
Holbrook, W. S., Mooney, W. D., and Christensen, N. I., 1992. Seismic velocity structure of the deep continental crust. In Fountain, D. M., Arculus, R., and Kay, R. (eds.), *Lower Continental Crust*. New York: Elsevier, pp. 1–43.
Johnston, J. E., and Christensen, N. I., 1995. Seismic anisotropy of shales. *Journal of Geophysical Research*, **100**, 5991–6003.
Mavko, G., Mukerji, T., and Dvorkin, J., 1998. *The Rock Physics Handbook*. Cambridge, MA: Cambridge University Press, pp. 289–303.
Rudnick, R. L., and Fountain, D. M., 1995. Nature and composition of the continental crust; a lower crustal perspective. *Reviews of Geophysics*, **33**, 267–309.

Cross-references

Deep Seismic Reflection and Refraction Profiling
Seismic Anisotropy
Seismic Imaging, Overview
Seismic, Reflectivity Method
Traveltime Tomography Using Controlled-Source Seismic Data
Vertical Seismic Profiling

SEISMIC QUIESCENCE AND ACTIVATION

Gennady Sobolev
Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, Russia

Definition

Seismic quiescence. Relative decrease in number of earthquakes or energy in any area of a seismic active region within a certain time interval in comparison with long-term observations in the same region.
Seismic activation. Relative increase in number of earthquakes or energy in any area of a seismic active region within a certain time interval in comparison with long-term observations in the same region.

Introduction

Both seismic quiescence and seismic activation before large earthquakes have been repeatedly described in the scientific literature. Large earthquakes are viewed as natural hazards, which may cause destruction of buildings, loss of life, and economic damage. Comparative size of different earthquakes is measured by their magnitude M (see *Earthquake, Magnitude*). The most destructive earthquakes fall in the range of magnitudes 7–9; and seismic energy released from an earthquake's source (see *Earthquakes, Energy*) lies in the range of 10^{22}–10^{25} ergs and the intensity of seismic shaking at the Earth surface (see *Earthquakes, Intensity*) exceeds the 7 level of the 12° scale.

Just the first condition of detection of quiescence or activation follows from the stated above: a necessity of long-term history of seismic observations. However, for the purpose of rigorously proven detection of such occurrences one should be sure that any seismic event, magnitude of which exceeded some minimum threshold, was not missed within the observation period. Just in the second half of the (twentieth) century development of seismic networks of observation equipped with calibrated seismographs lead to creation of instrumental catalogs of earthquakes what enabled to judge objectively about appearance of seismic quiescence and activation.

The first sufficiently founded reports of appearance of seismic quiescence before large earthquakes were published at the beginning of 1960s in Japanese and were practically unknown outside Japan. Then, in 1969, Mogi, on the basis of a visual analysis of seismicity maps, formulated a hypothesis that a seismic quiescence may be a precursor of a large earthquake (Mogi, 1979). At the end of the 80th year, Wyss and Habermann studied instrumental seismic catalogs of a number of regions of the Earth and determined basic rules of a formal detection of a seismic quiescence (Wyss and Habermann, 1988). In brief, they may be summarized as follows: (1) Evaluation of homogeneity of a seismic catalog for the analyzed time interval, including magnitude calculations, taking into consideration changes in methods of determination of coordinates and depth of seismic events. (2) Determination of a minimum magnitude of earthquakes being recorded without any omissions. (3) Removal of groups and aftershocks in order to analyze the so-called background seismicity. (4) Quantitative evaluation of size and significance of an anomaly, whereas significance shall be considered a statistically proven difference of an anomaly from random variations of the background seismicity. (5) Quantitative determination of the beginning of an anomaly. (6) Evaluation of sizes of an anomalous region. Wyss and Habermann (1988) determined that numerous events of seismic quiescence described in the literature are explained by changes of a minimum representative magnitude that were not taken into account, owing to the expansion of a seismic network or development of information processing technique. However, there remain dozens of published events of seismic quiescence that may not be explained by any omissions of technical nature (artifacts).

Description of certain events of seismic quiescence exceeds the limits of this article. We will mention only as an example two widely discussed events. Ohtake et al. (1977) studied the seismicity along the Pacific Ocean coast of Mexico and found that after the middle of 1973, the seismicity rate in the region with the linear dimensions of 100 by 200 km around coordinates of (16.5°N, 96.5°W) declined sharply what was interpreted as a seismic quiescence. The Oaxaca earthquake of November 29, 1978 with M = 7.6 occurred in this region. Kisslinger (1988) found a seismic quiescence before an Andreanof Islands earthquake of May 7, 1986 with M = 8. Our task does not include any evaluation of authenticity and significance of these and other events, especially because it is not correctly to do it without the authors' participation. We will

note only that about 100 authors of scientific works published their investigations of seismic quiescence in regions with different geological and tectonic structure. It may be assumed that such phenomenon exists objectively in the nature and sometimes arises before a large earthquake. Since a seismic process to a great extent is self-similar in a wide range of magnitudes, a seismic quiescence most likely arises before a relative weak earthquake as well.

Main phases of seismic quiescence and activation

On the basis of thousands of observations of seismicity in various regions of the world, there exists a generally accepted opinion among the seismologists that the large earthquake arises after a long-term (dozens of years) increase in seismicity in a relative region connected with a gradual growth of tectonic stresses. The latter, in their turn, arise at the joints of the earth's crustal plates moving with a different rate. Let us name this period of increase in seismicity as a *phase of long-term seismic activation* (phase I in the Figure 1). Spatial dimensions of a region of long-term activation before the large earthquake (M = 7 to 8) may exceed 1,000 km.

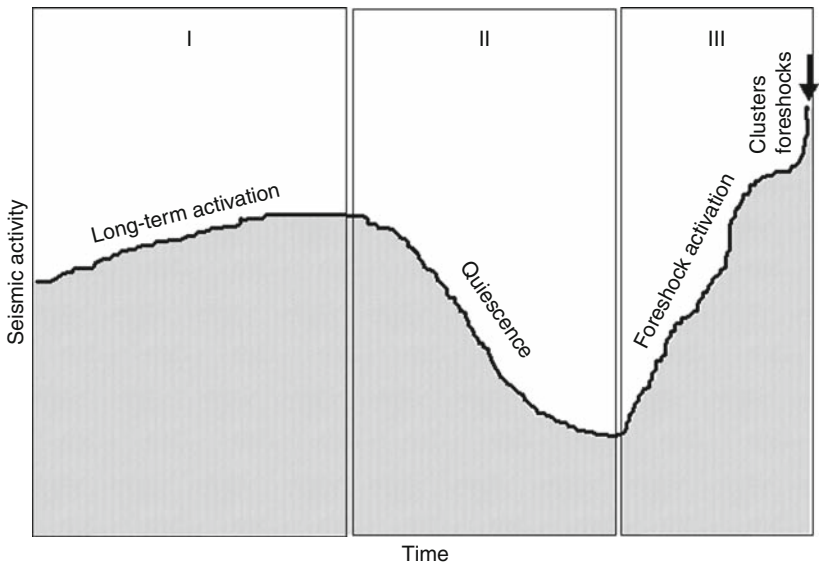
On the background of such process, in some area of the region of long-term activation a mean seismicity rate decreases what, as a rule, results in decrease in rate of release of the seismic energy. A researcher has a task to detect such decrease in the seismic activation and to prove its significance at a quantitative level, that is, to detect the seismic quiescence. All the suggested methods are based somehow or other on an analysis of statistics of earthquakes that at today's level of development of seismological networks of observations includes events contained in catalogs, the magnitude of which is from 3 to 4 units less in comparison with the magnitude of the large earthquake. This statistics includes at best the first thousands of events within a period of some years what makes the task of quiescence detection difficult and not always solvable. Below we will describe in brief two methods that were analyzed to reveal a seismic quiescence before several earthquakes both *a retrospective one* (after the large earthquake) and *a prospective one* (before the large earthquake).

In the Z-value method, there is calculated the standard deviation Z, to estimate the significance of the rate change,

$$Z = \frac{m_1 - m_2}{\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{1/2}},$$

where m is the mean rate, S the variance, and n the number of events in the first and second period to be compared (Wyss and Habermann, 1988).

For the purpose of significance evaluation there is applied the normal distribution law, but when having a rather large statistics of earthquakes the method gives plausible evaluations by other distribution law as well. The more the Z value is the more the difference between a rate of a seismic flow in the area of a supposed seismic quiescence and in comparison with a long-term rate. The calculations are carried out by the enumeration of nodes of the geographical grid and a changing number of earthquakes located near each node. It results the averaged Z values are presented in the form of maps in the researched seismic active region. The area of a seismic quiescence is determined according to high values of Z isolines above the given rate of statistic significance. The experience showed that the more reliable results appeared in the task if there were removed swarms and aftershocks of previous earthquakes from the catalog.



Seismic Quiescence and Activation, Figure 1 Main phases of seismic quiescence and activation.

The RTL method uses three functions to measure the state of seismicity at a given location as a function of time (Sobolev, 2001).

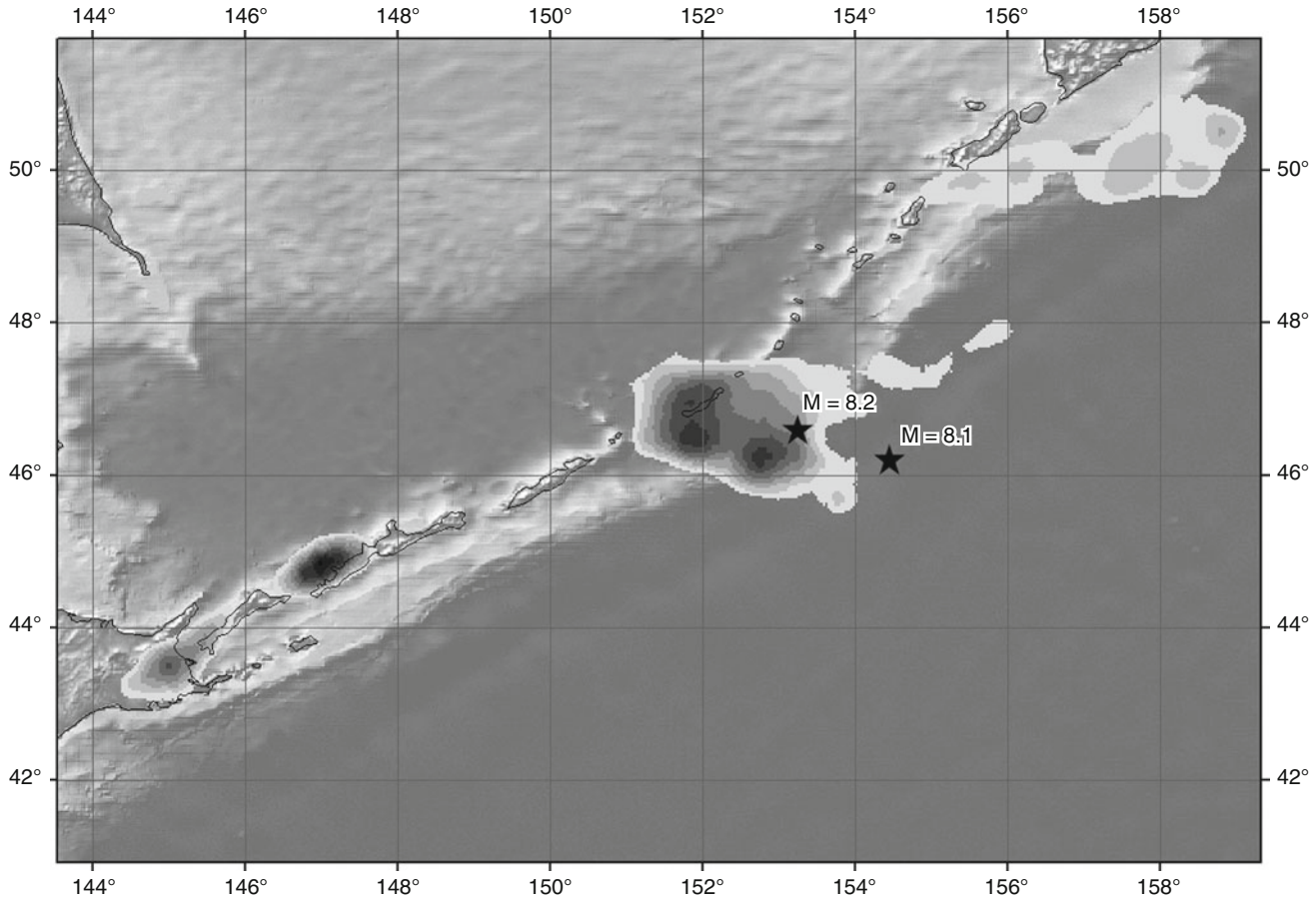
$$R(x,y,z,t) = [\Sigma \exp(-r_i/r_0)] - R_{\text{ltr}},$$
$$T(x,y,z,t) = [\Sigma \exp(-(t-t_i)/t_0)] - T_{\text{ltr}},$$
$$L(x,y,z,t) = [\Sigma \exp (l_i/r_i)^p] - L_{\text{ltr}}.$$

In these formulas, x, y, z, and t are the coordinates, the depth and the analysis time, respectively. r_i is the epicentral distance from the location selected for analyses, t_i is the occurrence time of the past seismic events, and l_i is the length of rupture estimated by empirical relation with the magnitude. The R_{ltr}, T_{ltr}, and L_{ltr} are the long-term averages of these functions. r_0 is a coefficient that characterizes the diminishing influence of more distant seismic events; t_0 is the coefficient characterizing the rate at which the preceding seismic events are "forgotten" as the time of analysis moves on; and p is the coefficient that characterizes the contribution of size of each preceding event. R, T, and L are dimensionless functions. They are normalized by their standard deviations. The product of the above three functions is calculated as the RTL-parameter, which describes the deviation from the background level of seismicity and is in units of the standard deviation, $\sigma_R\sigma_T\sigma_L$. The RTL-parameter decrease in relation to the background level means a seismic quiescence, and a structure of isolines of such parameter on the map indicates to the size of a quiescence area. In the RTL method, the outgoing catalog is cleaned of the aftershocks but the swarms are not removed.

Both methods (Z-value and RTL) require that the duration of period, which is included in calculations of the rate of background seismicity, must be by an order more than the supposed duration of the seismic quiescence. The experience of application of the described and other methods of revealing of the seismic quiescence in different tectonic conditions has shown that the duration of a quiescence period before the large earthquakes amounts to several years, that is, by an order less than the phase of the long-term activation (Figure 1). The area of significant values of the quiescence has linear dimensions about 300 km. In the most published works, the seismic quiescence was determined by the application of only one method. The exception, apparently, is a work by Wyss et al., in which a situation before the large earthquakes on the Sakhalin Island was analyzed by using of one and the same regional seismic catalog but by the application of both methods mentioned above: Z value and RTL (Wyss et al., 2004). The duration of quiescence anomalies before the Neftegorsk earthquake of May 27, 1995 with the M = 7.1, amounted according to results of both methods to 2.7 years. The size of a quiescence region was determined according to Z value in the form of a rectangle with dimensions of 200 × 600 km, and according to the RTL data – in the form of a circle with the radius up to 400 km. The epicenter of the earthquake lied at the periphery of anomalies in both determinations. Before the Uglegorsk earthquake of August 4, 2000, M = 6.8, the duration of quiescence was according to Z value 2.5 years and according to the RTL 3.0 years. The dimensions of the anomalous regions were determined in the form of a circle with the radius up to 165 and 200 km, respectively. The epicenter of the earthquake was also at the lateral part of quiescence area. The examination conducted with random catalog has shown that the probability that the anomaly occurred by chance before the Neftegorsk earthquake did not exceed 1% and before the Uglegorsk earthquake – 2% according to the results of both methods.

In spite of available facts of the seismic quiescence before some large earthquakes, the practical application of such anomaly for the forecast of earthquakes remains undecided. Authors of publications on the subject of seismic quiescence do not cite any data on a number of false alarms when no large earthquake occurred in the region of quiescence. According to our evaluations, the number of false alarms exceeds by several times the number of forecasts proven to be correct (even by a retrospective analysis). There is no statistic evaluation of a real *prospective* forecast with the use of such anomaly indeed. Let us illustrate the above stated by Figure 2, where the areas of seismic quiescence in the region of the Kurile arch are shown, which were detected before the Simushir earthquake of November 15, 2006, M = 8.2. In this case the RTL method was applied, but it is not a matter of principle. The dark spots mean the areas of seismic quiescence. The main one of them with linear dimensions of about 200 km is located in on the territory with coordinates of 46°–48°N – 151°–154°E; the epicenter of the main shock is located at its periphery, and the another major earthquake of January 13, 2007, M = 8.1 – outside the anomalous region. Several other anomalies of a lower size are shown in the figure: the territory with coordinates (45°N, 147°E), (43°N, 145°E), and (50°N, 158°E). Any large earthquakes in these regions did not occur up to now.

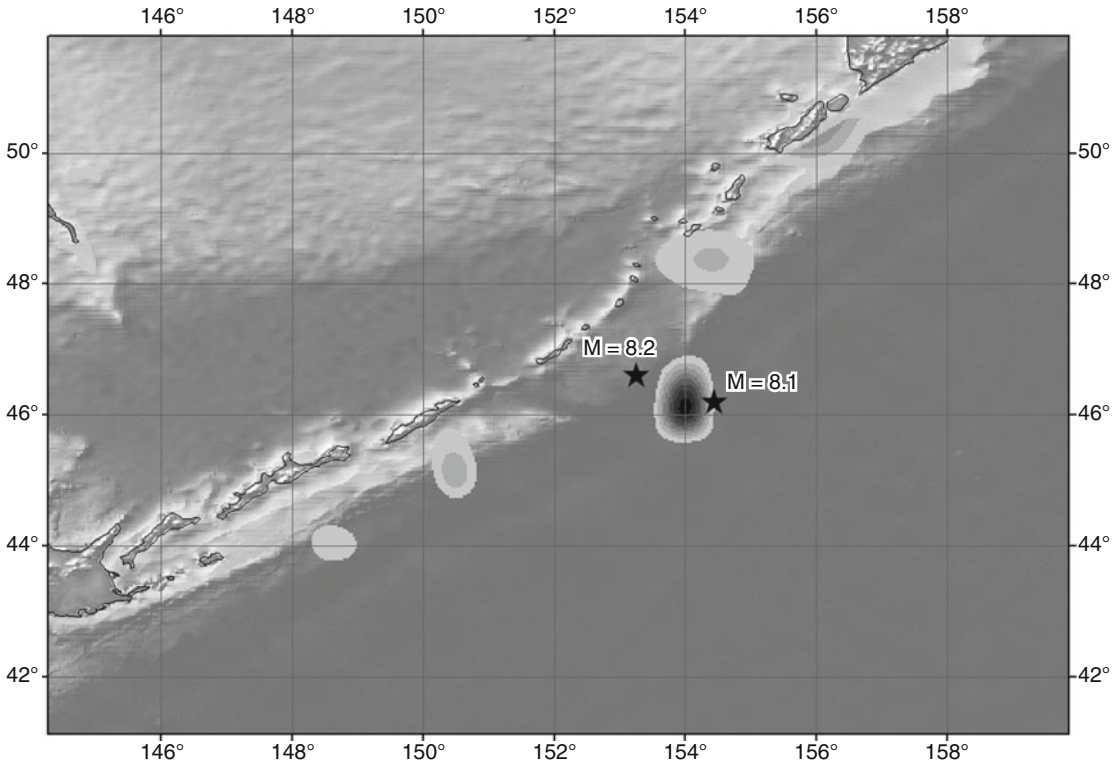
It follows from the experience of investigations of the seismic quiescence that a large earthquake seldom occurs during the most significant rate of quiescence. It arises more often after the rate of seismic activity has increased again and even exceeded that which has been observed before the beginning of the quiescence. Let us name the time interval from minimal values of activity at the stage of quiescence to the moment of the large earthquake as a phase of foreshock activation (phase III on Figure 1). Its duration has a wide range: from several days to the dozens of months. The areal of foreshock activation in general coincides with the area of aftershocks (see *Earthquake, Aftershocks*) and its size is estimated in hundreds of kilometers. These properties of the phase of foreshock activation bring to the understanding of the term "foreshock" in the broad sense. At the final step of development of such phase there may occur seismic events, the locations of which coincide practically with the source of the



Seismic Quiescence and Activation, Figure 2 Areas of seismic quiescence before the Simushir earthquake.

following large earthquake. They are traditionally called foreshocks and should be understood as foreshocks in the narrow sense (see *Earthquake, Foreshocks*). Usually they arise several hours before a large earthquake, reflecting apparently the nucleation of a rupture. This phenomenon is of a great interest for a short-term forecast, but the practical application thereof has some difficulties because no reliable criteria are found till today in order to differentiate foreshocks from usual earthquakes. In exceptional events when a number of events occurring during several hours and at one and the same place amounts to dozens, they may be interpreted presumably as foreshocks of a future large earthquake and serve as a ground for a short-term forecast. The striking example is a prediction of the Haicheng earthquake of February 4, 1975, $M = 7.3$ (Ma Zongjin et al., 1989). The foreshocks determined after the large earthquake (retrospectively) are described in a great number of publications and the description thereof is beyond the scope of this article.

Returning to the phase of foreshock activation (Figure 1), we cite the example of its detection before the Simushir earthquake mentioned above as of November 15, 2006, $M = 8.2$. The RTL method developed for the detection of a seismic quiescence allows also to identify the foreshock activation as a period of increase in the seismic activity after the quiescence. The situation showed in Figure 3 is a continuation in time of that which appeared in this area of the Kurile arch during the seismic quiescence (Figure 2). The development of the stage of seismic quiescence from the level of a long-term background to the lowest values of the seismic activity lasted in this event 1.5 years and the period of recovery of the stage of activity to the level of a long-term background (foreshock activation) lasted 1.5 years as well. The linear dimensions of foreshock activation amounted to about 100 km (Figure 3). It was located in the area of aftershocks between the main shock with the $M = 8.2$ and the strongest aftershock, $M = 8.1$. Also, like at a seismic quiescence, within the region as showed on the map (Figure 3) during the development of the main anomaly there was selected a number of anomalies with a lower intensity: on the territory of coordinates (44°N, 149°E); (45°N, 150.5°E); (48.5°N, 154.5°E.). In these places, no large seismic events occurred up to now. Thus, the foreshock activation, like the seismic quiescence, is not of high reliability in the prognostic aspect because there exist anomalies qualified as false.



Seismic Quiescence and Activation, Figure 3 Areas of foreshock activation before the Simushir earthquake.

Physical mechanisms leading to seismic quiescence and activation

Let us consider possible physical mechanisms, which bring to occurrences of the seismic quiescence and of the foreshock activation. There is no unified explanation at present and it is expedient to discuss various hypotheses. We would like to note at first that except for the three phases discussed above and the occurrence of foreshocks in the narrow sense, other anomalies may also appear sometimes. Scholz selected additional phases: doughnut pattern was observed in the period of development of the seismic quiescence around its external limits, and seismic silence was noticed just prior to the main event (Scholz, 1990).

Let us consider in a more detailed manner the hypotheses of nature of the stages of a seismic quiescence and of a foreshock activation. Ma Zongjin et al., investigated the seismic patterns before nine large earthquakes in China (Ma Zongjin et al., 1989). The authors did not select specially occurrences of a seismic quiescence or activation. They did describe mainly a migration of sources of seismic events for several years before large events in the regions adjacent to their epicenters with the dimensions of about first 100 km. Actual data cited by them are of interest as applied to a problem of physical nature of occurrences discussed in this article. A migration of sources was considered, mainly with the magnitudes 3–6, while the large earthquakes had the magnitudes of above 7. The regions round the epicenters of large events were separated in three parts: a hypocentral area A, which included a hypocenter of a large event; area of aftershocks B; and external area C.

The following main characteristics of migration were selected: During 3–10 years, till the moment of the large earthquake, two phases were varied in time. In the first one the events occurred mainly in the area C, while the areas A and B were characterized as relatively quiet. During the second phase which finished in the large earthquake, the activity was detected in the areas A and B, while the seismic activity in the area C was reduced. As the second phase developed, the activity in the areas A and B continued to last up to several months or days in relation to the moments of the large earthquake; at the end of such series the activity either accumulated inside the hypocentral area A or continued to last in the area B, while the area A was calm. We should note that the authors did not cite any quantitative evaluations of the seismicity rate, thus conclusions may be made based on a qualitative manner only. In summarization, there may be made a conclusion that the process of development of a large earthquake may occur according to various scripts, though in described events it (is) related to continental events only.

The seismic quiescence may be a consequence of increase in the strength of rocks within the seismically active area. In the dilatancy-diffusion model the increase in the strength is explained by laboratory experiments, being a consequence of appearance of open microcracks

and a relative drying of the rocks. However, it is not explained how such process can develop in the heterogeneous lithosphere in the spatial regions with the linear dimension of above 100 km.

The quiescence occurs also when the acting stresses decrease. This can be a consequence of development of an unstable deformation in the source of a future large earthquake what constitutes one of the corner stones of the model of avalanche-like fracturing. Expenditure of the accumulated potential energy brings to a decrease in stress both in the source and in the surrounding space. This explanation leads to expect an increase in seismic activity in the hypocentral area of a future large earthquake and a simultaneous occurrence of a seismic quiescence in the external area. This cannot always be observed. But the situation that the activation in the source occurs at the level of seismic events with small magnitudes, which are not registered by the available seismological network, cannot be excluded.

Decrease in the tectonic stress in some area of lithosphere can be a consequence of the motion of the neighboring blocks with a different rate. If one of the blocks stops because of a strong asperity at its tectonic boundaries with any neighboring blocks, then the stress will be accumulated at points of an asperity, while in the middle of the stopped block it will relax gradually. The activation will start at boundaries of the stopped block and then it will migrate to central parts when the asperity is destroyed.

As a modification of such assumption is the situation when the stress increases and brings to the seismicity in the stronger blocks, and the block between them is relatively discharged. The quiescence occurring in the discharged block in such situation causes an effect of "false forerunners" as no large earthquake follows after them.

Kanamori suggested an explanation of different phases of seismicity based on a model of existence of a strong inclusion at a fault being heterogeneous in its strength (Kanamori, 1981). When the stress increases gradually, the less strong inclusions disintegrate sequentially what brings to an accelerated growth of the stress at points being not disintegrated. A background seismicity is observed at the low stress. As the stress increases, there occurs the doughnut pattern – a mass destruction of rocks around the strong inclusion. The latter has a seismic quiescence. When the stress approaches to a critical level, there occurs a destruction of subunits of the strong inclusion which characterizes the phase of foreshock activation. In this hypothesis, the facts of existence of the seismic quiescence at distances that significantly exceed the rupture sizes when a large earthquake occurs are not explained.

The reason of seismic quiescence can be a change of orientation of the tensor of the current stress due to development of a creep at the fault where a future earthquake will occur. The existing fractures, which get into new conditions, require the time for unstable development (kinetics of destruction). It is not proven that this, undoubtedly, existing phenomenon can cause a quiescence at distances that significantly exceed the length of a rupture of a future earthquake.

Another reason can be a transformation of the medium surrounding the source to the quasi-plastic state, for example, when the temperature or the hydrostatic compression increases. Then the process of fracturing will occur at a lower scale level which is beyond the registration of earthquakes by the seismic network. There remains a question what are the physical reasons of increase of the temperature or the compression in the lithosphere or the earth's crust within a relatively short time period (years).

It is necessary to mention another effect, which influences on the fracturing in the geological medium. Laboratory and field observations prove that an increase or a decrease in the degree of water saturation of rocks results in a significant acceleration or retardation of occurrence of the seismic events (seismicity, triggered/induced). Further, if the rate of the seismic activity increases due to that, then it results in a relative increase in the number of events with the relatively small magnitudes and in a decrease in the number of big events. The seismic quiescence may occur: (a) if a degree of water saturation decreases, (b) if under the influence of the increased water saturation the seismicity transforms to the level of small events being not registered by the seismic network. The activation is a direct consequence of the increased water saturation and may appear also in the form of swarms.

It follows from the most available experimental facts that the foreshock activation develops in the epicentral region. Its size is less than the area of seismic quiescence, and the centers of anomalous areas of these two phenomena do not coincide as a rule. There is a ground to suppose that the physics of foreshock activation is connected with the development of unstable deformation, which is localized under the laws of mechanics in the zone of mainly two-dimensional extension. The complexity of fault systems, which display a fractal geometry, causes a successive occurrence of several zones of unstable deformation.

At the final period of foreshock activation, there often occur clusters of seismic events, that is, groups of events, the distances between the hypocenters and the times between the successive events of which are less than mean values of the background seismicity. The occurrence of clusters at the phase of foreshock activation can be explained by two reasons. First, they occur by chance because of the increase in rate of formation of events in the narrow zone of unstable deformation. Second, when the spatial density of accumulated active faults exceeds the critical level, there arise stress interactions among neighboring faults with the formation of the faults of a larger length. In the latter case, such effect must appear in the increase of middle magnitudes of seismic events, that is, in the decrease of the b value.

Summary

It follows from the above-mentioned hypotheses that there exist various physical mechanisms of occurrence both of the seismic quiescence and of the foreshock activation. For the purpose of more fundamental understanding of

such phenomena, the additional laboratory and fieldworks are required. In our opinion, special attention should be paid to the following directions.

To investigate on the same catalogs of earthquakes both the quiescence and the activation in a complex, paying attention to their distribution in space, time, and magnitudes.

To compare, where possible, the seismic patterns with the field of deformations being estimated based on the data of satellite geodesy.

To compare, where possible, the seismic patterns with the data of deep geoelectrical and hydrogeological investigations with the purpose of better understanding of the role of water.

Bibliography

Kanamori, H., 1981. The nature of seismicity patterns before large earthquakes. *Earthquake prediction. International Review.* Washington, DC: American Geophysical Union, pp. 1–19.

Kisslinger, C., 1988. An experiment in earthquake prediction and the 7 May 1986 Andreanof Islands earthquake. *Bulletin of the Seismological Society of America*, **78**, 218–229.

Ma, Z. J., Fu, Z. X., Zhang, Y. Z., Wang, C. G., Zhang, G. M., and Liu, D. F., 1989. *Earthquake Prediction, Nine Major Earthquake in China*. New York: Seismological Press, Springer, 332 pp.

Mogi, K., 1979. Two kinds of seismic gap. *Pure and Applied Geophysics*, **117**, 1172–1186.

Ohtake, M., Matumoto, T., and Latham, G. V., 1977. Seismicity gap near Oaxaca, Southern Mexico, as a probable precursor to a large earthquake. *Pure and Applied Geophysics*, **115**, 375–385.

Scholz, C. H., 1990. *The Mechanics of Earthquakes and Faulting*. Cambridge: Cambridge University Press, 439 pp.

Sobolev, G., 2001. The examples of earthquake preparation in Kamchatka and Japan. *Tectonophysics*, **338**, 269–279.

Wyss, M., and Habermann, R. E., 1988. Precursory Seismic quiescence. *Pure and Applied Geophysics*, **126**, 319–332.

Wyss, M., Sobolev, G., and Clippard, J. D., 2004. Seismic quiescence precursors to two M7 Earthquakes on Sakhalin island, measured by two methods. *Earth Planets Space*, **56**, 725–740.

Cross-references

Artificial Water Reservoir Triggered Earthquakes
Earthquake, Aftershocks
Earthquake, Foreshocks
Earthquake, Magnitude
Earthquakes, Energy
Earthquakes, Intensity

SEISMIC SEICHES

Art McGarr
U.S. Geological Survey, Menlo Park, CA, USA

Definition

Seismic seiche is a term first used by Kvale (1955) to discuss oscillations of lake levels in Norway and England caused by the Assam earthquake of August 15, 1950. This definition has since been generalized to apply to standing waves set up in closed, or partially closed, bodies of water including rivers, shipping channels, lakes, swimming pools and tanks due to the passage of seismic waves from an earthquake.

The first published mention of seismic seiches is thought to be reports of those observed throughout much of Europe due to the great earthquake at Lisbon, Portugal in 1755 (Wilson, 1953; Richter, 1958). In addition to the Lisbon and Assam earthquakes, seismic seiches at teleseismic distances have been observed for many other large earthquakes including the 1964 Alaska (McGarr and Vorhis, 1968) and the 2002 Denali, Alaska, an earthquake that caused damaging seiches in Lake Union, Seattle, Washington at an epicentral distance of 2,400 km (Barberopoulou et al., 2004).

Kvale (1955) showed that seismic surface waves from the Assam earthquake were the most probable cause of the seiches observed in Norway and England at that time. Moreover, he concluded that the natural period of a basin must be matched by the periods of the passing seismic surface waves. Motivated by observations reported by Donn (1964) of a seiche generated in a channel near Freeport, Texas, at an epicentral distance of about 5,040 km from the 1964 Alaska earthquake, McGarr (1965) developed a relation between the ground motion of seismic waves and the resulting seiche. The passing seismic wave exerts a horizontal acceleration on a closed or partially closed body of water, which can be idealized as a long channel of uniform depth. This causes a seiche, composed of standing water waves whose periods depend on the dimensions of the channel. The amplitude of the seiche is a function of channel depth, the amplitudes of the horizontal accelerations of the passing seismic waves, and the extent to which the periods of the seismic waves match those of standing water waves. The gravest mode of the standing waves has a period given by $T = 2L/\sqrt{gH}$, where T is the period in seconds, L is the channel width in meters, H is the channel depth in meters, and g is gravity. For instance, if L = 100 m and H = 10 m, then the period of the gravest seiche mode is 20 seconds, which tends to be the period where surface waves show maximum amplitudes at teleseismic distances. Any factor that enhances the amplitudes of surface waves, such as basins containing low-velocity sediments, tends to result in greater production of observed seiches from a given earthquake (McGarr and Vorhis, 1968; Barberopoulou et al., 2004).

Bibliography

Barberopoulou, A., Qamar, A., Pratt, T. L., Creager, K., and Steele, W. P., 2004. Local amplification of seismic waves from the Denali earthquake and damaging seiches in Lake Union, Seattle, Washington. *Geophysical Research Letters*, **31**, L03607, doi:10.1029/2003GL018569.

Donn, W. L., 1964. Alaska earthquake of 27 March, 1964: Remote seiche stimulation. *Science*, **146**, 261–262.

Kvale, A., 1955. Seismic seiches in Norway and England during the Assam earthquake of August 15, 1950. *Bulletin Seismological Society of America*, **45**, 93–113.

McGarr, A., 1965. Excitation of seiches in channels by seismic waves. *Journal of Geophysical Research*, **70**, 847–854.
McGarr, A., and Vorhis, R. C., 1968. Seismic seiches from the March 1964 Alaska earthquake. *U.S. Geological Survey Professional Paper*, **544E**, E1–E43.
Richter, C. F., 1958. *Elementary Seismology*. San Francisco: WH Freeman, 768 pp.
Wilson, B. W., 1953. Coastal seiches, pt. 1 of Oscillations of the sea and the phenomenon of range. The Dock and Harbour Authority [London], pp. 41–45.

SEISMIC SIGNALS IN WELL OBSERVATIONS: PRE, CO, POST

R. K. Chadha
National Geophysical Research Institute, Hyderabad, Andhra Pradesh, India

Synonyms

Abnormal water level changes; Hydrological earthquake precursors

Definition

Seismic signals in wells – rise or fall in water level fluctuations in bore wells due to earthquakes. Pre, co, and post means abnormal fluctuations prior, during, and after the occurrence of earthquake.

Introduction

Water levels in wells are found to be affected by local and distant earthquakes in addition to other phenomena like, rainfall, atmospheric pressure, and earth tides. These anomalous changes that could be pre, co, and post earthquake are believed to reflect pore pressure changes related to the redistribution of stress in the near and far fields of dislocation sources, but in general are poorly understood because of inadequate data. Numerous reports of the well-documented cases of this phenomenon have appeared from China, Japan, Russia, the USA, India, Taiwan, and other countries during the last 30 years (Wakita, 1975; Roeloffs, 1988, 1996, 1998, 2003; Liu et al., 1989, 2006; King et al., 1999, 2000; Rudnicki et al., 1993; Koizumi et al., 1996, 1999; Quilty and Roeloffs, 1997; Chadha et al., 1997, 2003, 2008; Gavrinlenko et al., 2000; Wen-Chi Lai et al., 2004; Akita and Matsumoto, 2004; Kitagawa et al., 2006; Sil, 2006a, b). Since the earthquake-related groundwater level changes were scattered and rarely observed by a sufficiently dense network of observations wells, such studies have not been adequately developed. Recently, some experiments have been taken up to drill bore wells in the seismically active region to continuously monitor water-level changes related to earthquakes.

Types of observed changes

Four types of changes in well water levels have been observed, which are related to earthquakes, namely, pre, co, post, and transient. While, the pre and post earthquake observations are mostly interpretative in nature and difficult to substantiate, coseismic and transient changes are well established. The most suitable wells for observing anomalous coseismic and transient changes are the ones that are connected to confined aquifers and show the effect of earth tides. The presence of tidal signals in well level data indicates that the well is sensitive to small strain changes in the connected rock formations and hence also should be sensitive to variations in local stress fields, and thus to earthquakes. Unconfined aquifers also show seismic signals in case of large magnitude earthquake at closer distances. Few typical examples of coseismic and transient changes are described below.

Coseismic seismic signals

Coseismic signals are generally observed in wells connected to confined aquifers and are located in the near field of earthquake source location. These are generally steplike changes coinciding with the occurrence of the earthquake. In case of shallow unconfined aquifers, coseismic oscillatory changes have also been reported.

The Chi-Chi earthquake of Mw 7.6 on September 21, 1999 is one of the well-documented events for earthquake-induced water level changes in wells. Based on the hourly digital record of the water levels, all changes were reported as coseismic or postseismic. Chia et al. (2001) described the details of these changes and compared the coseismic changes with the distances between the earthquake fault and observation wells. Wang et al. (2001) discussed the distribution of the coseismic changes and attributed them to liquefaction. Wen-Chi Lai et al. (2004) compared these changes with geological setting and seismic ground motion. They showed that in the nearby Choshui River alluvial fan area, the groundwater levels coseismically rose and those amplitudes increased as the ground acceleration and hydraulic conductivity became larger. In the slope area near the earthquake fault, the water level coseismically dropped and those amplitudes increased as the ground acceleration became larger.

Akita and Matsumoto (2004) reported coseismic changes in 29 wells associated with M 8.0 Tokachi-oki earthquake in 2003 in Japan. These changes were observed as increase in groundwater in wells located in the southeast part of Hokkaido. The maximum increase was 430 cm. In other area of Hokkaido, coseismic decreases were observed, the maximum being 59 cm. These anomalous changes were explained in terms of poroelastic response to the volumetric strain after the Tokachi-oki earthquake. The strain sensitivities determined by coseismic responses in the groundwater level were found to be consistent with those estimated by M_2 tidal strain in few wells.

Matsumoto et al. (2003) studied hydrological response at Haibara well to earthquakes in central Japan. They reported 28 coseismic changes during the period from 1981 to 1997 and obtained a relationship $M \geq 2.45 \log_{10}$

Dis + 0.45 which holds good for 26 events. Most of the water level changes in Haibara well could not be explained by poroelastic response of the water level to coseismic volumetric strain steps in confined aquifers, because the sizes and directions of the coseismic water level changes correlate poorly with the sizes and directions of the volumetric strain steps. Ground motion due to seismic waves was put forward as a more probable reason for coseismic drops in water level. Few preseismic or inter-earthquake changes were also observed by them in the residual water level data at Haibara well, which was interpreted to be related to local aseismic crustal deformation.

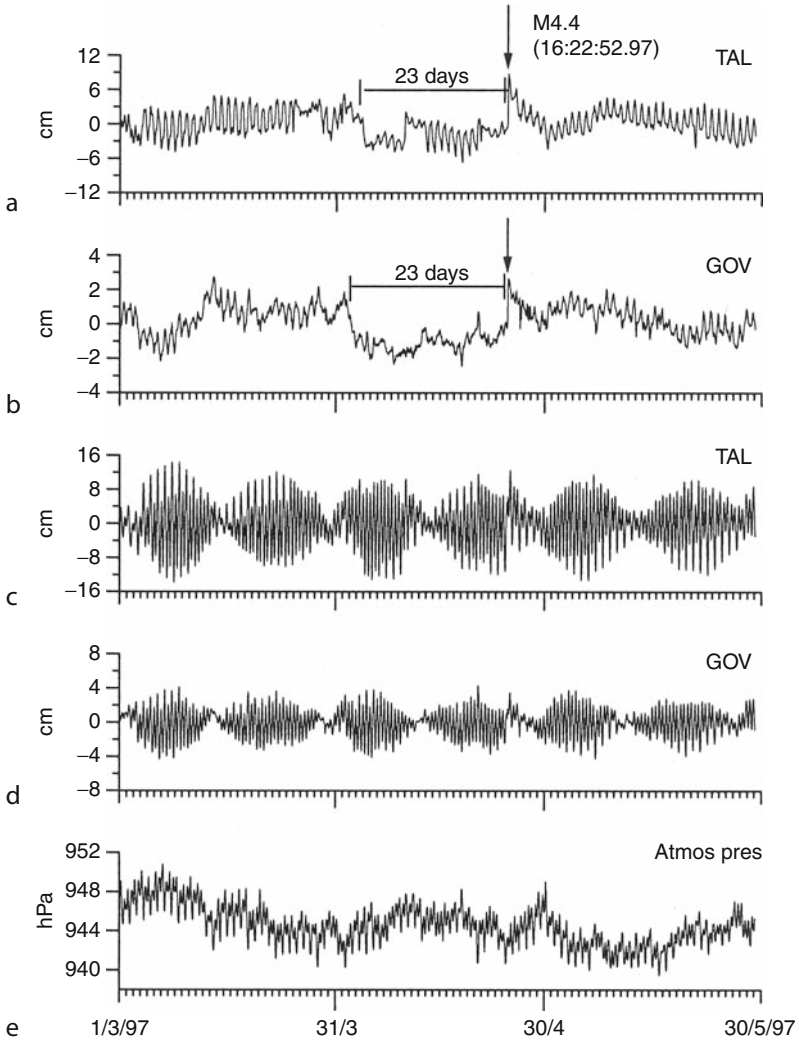
Along the San Andreas fault in the Parkfield area, water levels at 17 wells were monitored continuously as a part of Parkfield earthquake prediction experiment. Rudnicki et al. (1993) analyzed data from a well located 460 m from the San Andreas fault near Parkfield, California during January 1989–July 1990. They demonstrated that recovery rates of water level changes due to fault creep can be used to infer the position of the slipping zone relative to the well. They applied this technique to five slip-related water level changes. The changes were all characterized by a rapid drop (less than 8 h) and a slow recovery (15–30 days). The recoveries modeled by pore fluid diffusion due to a plane strain dislocation in porous, fluid-saturated elastic provided satisfactory agreement with the observations. Calculations were also done for limiting cases of both an impermeable and a permeable fault. They showed that permeability of the fault has little effect on inferences about the slip, but the inferred (horizontal) diffusivity differs for the two cases, 0.15 m^2/s for the impermeable and 0.06 m^2/s for the permeable. Later, Roeloffs (1998) analyzed persistent water level changes data at a well near Parkfield for anomalies caused by local and distant earthquakes.

Chadha et al. (2003) reported four cases of coseismic changes in well around Koyna-Warna region in western India. Koyna region is known to be site of the world's largest Reservoir Triggered Earthquake of M 6.3 in 1967. For the last 5 decades, earthquakes of $M \geq 5.0$ continue to occur in a small source volume of the dimension $30 \times 15 \times 10$ km^3 in this region. To comprehend the cause-and-effect relationship between local earthquakes and water level changes, 21 boreholes were drilled surrounding the seismic source volume. Out of the 21 boreholes, 10 were found to be connected to confined aquifers and showed strong tidal signals. Analysis of data from these wells from 1997 to 2003 revealed four cases of coseismic changes associated with earthquakes of $M \geq 4.2$. Figure 1 shows a coseismic steplike increase in two wells in the Koyna region, which is preceded by precursory drop in water levels prior to the M 4.4 earthquake. While coseismic increase of 2–7 cm in water levels was observed for earthquakes of M 4.3–4.7 in three cases, a troughlike decrease up to 8 cm was observed at seven wells for an M 5.2 earthquake. All these earthquakes occurred within the network of wells drilled for the study. From their studies, Chadha et al. (2003) concluded that the magnitude of the earthquake and epicentral distance to the wells are two vital parameters for recording the hydrological anomalies in well water levels. In Koyna, coseismic anomalies were recorded in wells for earthquakes of $M \geq 4.3$ located up to 24 km distance. Although there were several earthquakes of $M < 4.0$ during the period of study, no anomalous changes were observed in well recordings. Also, few precursory anomalies were also interpreted for local earthquakes that showed coseismic changes.

Using data of M 4.4 earthquake in Koyna, Chadha et al. (2005) tested the hypothesis that "well level fluctuations respond to changes in crustal volume strain that is induced by an earthquake in the form of step like coseismic change." Using Okada's (1992) formulation, they calculated the static volumetric strain at the surface of a homogeneous half space to see whether the coseismic steps observed for this earthquake agree with the observations of Wakita (1975) that the water level in a well shall rise or fall based on the location of the well with reference to the disposition of the fault plane. The results showed that expanding and contracting zones representing the static volumetric strain field agree with the observations of well level changes indicating that well-aquifer system indeed has the potential to reflect coseismic volumetric strain changes, similar as with the tidal fluctuation in volume strain. Further, using the strain sensitivities obtained from tidal analysis, they showed that there was misfit in amplitudes of the observed and calculated coseismic steps. Based on this study, they concluded that coseismic strains may be a function of site effects controlled by local heterogeneity in geological structures. Thus, simple elastic models cannot explain the amplitudes of hydrological anomalies wholly. Huang et al. (1995) have earlier suggested a nonlinear response of water levels to coseismic strains due to local heterogeneities. Earlier, Grecksch et al. (1999) have also observed coseismic well level steps to be higher than predicted from strain sensitivities.

Transient seismic signals

Seismic waves from distant earthquakes can cause changes in well water levels at great distances due to the passing of seismic waves (Rexin et al., 1962). There are several reports of such changes from different parts of the world, following the Sumatra earthquake of Mw 9.3 on December 26, 2004 (Kitagawa et al., 2006; Sil, 2006; Sil and Freymueller, 2006; Wang and Manga, 2006; Chadha et al., 2008). Earthquake-induced water level changes at distant locations were also reported after the Denali earthquake (Brodsky et al., 2003; Harp et al., 2003; Stricherz, 2003; Cassidy and Rogers, 2004; Kayen et al., 2004; Sil, 2006). It is thought that dynamic oscillation in crustal strain in an aquifer due to a seismic wave mainly caused dynamic oscillation in the groundwater level. Seismic oscillations, due primarily to surface waves from distant events, occur in some wells tapping highly transmissive aquifers (Liu et al., 1989; Liu et al., 2006). Earlier, several workers have shown that anomalous



Seismic Signals in Well Observations: Pre, Co, Post, Figure 1 Records of air pressure and well water level in GOV and TAL wells, during March 1–May 30, 1997. (**a**) and (**b**) residual water levels after removing earth tides and atmospheric pressure, (**c**) and (**d**) raw data after removing the trend. Spring and neap tides are clearly seen, (**e**) Atmospheric pressure. Arrow shows origin time of the earthquake on April 25, 1997. Precursory decrease 23 days prior to M 4.4 earthquake is shown by a horizontal line.

well-water level changes occur in response to deformation of the connected aquifers because of seismic waves (Cooper et al., 1965, 1967; Liu et al., 1989), fault creep (Wesson, 1981; Roeloffs, 1996, 1998), tidal strain (Bredehoeft, 1967; Van der Kamp and Gale, 1983), or atmospheric loading (Rojstaczer, 1988). Cooper et al. (1965) and Kunugi et al. (2000) showed that the amplitude of the oscillation in groundwater level is enhanced in a particular period due to the characteristics of the well-aquifer system.

M 9.1 Sumatra earthquake of 2004 and transient seismic signals

Kitagawa et al. (2006) reported water level changes in wells due to Sumatra earthquake of 2004 at a distance of more than 5,000 km in Japan. At 38 of the 45 observation stations, these changes were recorded. Ishii-type borehole strain instruments installed in 10 of the observation stations also recorded changes in crustal strains. Most of these changes in crustal strains and groundwater levels were dynamic oscillations due to a seismic wave. Sil and Freymueller (2006) reported anomalous changes in wells in Fairbanks, Alaska due to the passing of seismic waves of Sumatra earthquake. Chadha et al. (2008) reported large water level changes in six bore wells in the Koyna region in western India, following the Great Sumatra earthquake of Mw 9.3 in December 26, 2004. This earthquake occurred at 00:58:50 coordinated universal time (UTC) off the coast of Sumatra located about 3,000 km from the Koyna region. The arrival time of the *P* wave recorded at the Koyna seismic station was 01:04:24.45 (UTC). The anomalous water level changes were observed at 01:15

(UTC) at all of these wells where the sampling rate was 15 min. No abnormal changes were noticed in the preceding sampling time at 01:00 (UTC), clearly indicating that these changes occurred due to the passage of seismic waves generated by the Sumatra earthquake. However, due to the large sampling interval, it was difficult to correlate the water level changes either to shear or surface waves. No local earthquake was recorded by the network during this period. The anomalous water level changes were of the order of 2–65 cm and showed either a spikelike or a steplike change. These types of changes are attributed to the dynamic strain induced by the passage of seismic waves, most probably long period surface waves.

Summary and conclusion

Earthquakes can cause changes in well water levels in near and far fields. This has been exemplified by several reported cases in the literature during the last 3 decades. Therefore, water level monitoring becomes a valuable tool for earthquake prediction research. Although, the coseismic and transient water level changes are well established, the greatest challenge is to identify precursory anomalous changes that can be identified prior to the occurrence of earthquakes. Very often, these precursory changes that represent the pre-earthquake phase are very small and therefore not very obvious in the raw data. This is due to the response of wells to large atmospheric pressure, significant precipitation, or earth tides, which may obscure these smaller earthquake-related changes. Sophisticated filtering techniques have to be developed to decipher the earthquake related anomalies from non-tectonic effects in the data. Definite patterns of water level anomalies have to be established, both site specific and global, so that predictive capabilities can be developed. This could be done with continuous monitoring of dedicated networks of wells around known seismically active regions in different countries.

At present, three categories of anomalies have been reported, namely, co- and preseismic, aseismic, and transient changes. In the first category, a steplike coseismic rise or fall has been observed. This type of coseismic steps is understood as sudden pore pressure changes related to an alteration in in-situ volume strain caused by the redistribution of stress in the brittle crust (Bodvarsson, 1970; Kuempel, 1992; Muir-Wood and King, 1993). Sometimes, these coseismic steps were found to be preceded by persistent water level drops prior to the earthquakes. According to Sadovsky et al. (1979) and Monakhov et al. (1983), the most common precursor is a water level drop of a few centimeters amplitude several days before the earthquake. Typically, the drop is beginning to recover when the earthquake occurs. This type of anomaly has been referred to as "rebound anomaly" (Igarashi et al., 1992) and is believed to be related to an increase of porosity and permeability due to fracturing, with the subsequent recovery attributable either to influx of fluid or to compression (Roeloffs, 1988). Another type of precursory water level drop that occurs much prior to coseismic steps, but recovers just before the earthquake. This type of water level drops could be due to aseismic creep along a fault plane where the earthquake stress is building up. The third type of precursory change occurs and recovers much before the occurrence of the earthquake. This type of water level drops and recovery is believed to be due to small slips on the hanging wall of a normal fault prior to occurrence of the main earthquake. All these precursory changes are established in the hindsight, that is, after the occurrence of earthquakes. Continuous data sets for long periods will help in establishing these kinds of anomalies.

Bibliography

Akita, F., and Matsumoto, N., 2004. Hydrological responses induced by the Tokachi-oki earthquake in 2003 at hot spring wells in Hokkaido. Japan. *Geophysical Research Letters*, **31**, L16603, doi:10.1029/2004GL020433.

Bodvarsson, G., 1970. Confined fluids as strain meters. *Journal of Geophysical Research*, **75**, 2711–2718.

Bredehoeft, J. D., 1967. Response of well-aquifer systems to earth-tide. *Journal of Geophysical Research*, **72**, 3075–3087.

Brodsky, E. E., Roeloffs, E., Woodcock, D., Gall, I., and Manga, M., 2003. A mechanism for sustained groundwater pressure changes induced by distant earthquakes. *Journal of Geophysical Research*, **108**, 2390.

Cassidy, J. F., and Rogers, G. C., 2004. The Mw 7.9 Denali fault earthquake of 3 November 2002: felt reports and unusual effects across western Canada. *Bulletin of the Seismological Society of America*, **94**, S53–S57.

Chadha, R. K., Gupta, H. K., Kuempel, H.-J., Mandal, P., Nageswara Rao, A., Kumar, N., Radhakrishna, I., Rastogi, B. K., Raju, I. P., Sarma, C. S. P., Satyamurthy, C., and Satyanarayana, H. V., 1997. Delineation of active faults, nucleation process and pore pressure measurements at Koyna (India). *Pure and Applied Geophysics*, **150**, 551–562.

Chadha, R. K., Pandey, A. P., and Kuempel, H. J., 2003. Search for earthquake precursors in well water levels in a localized seismically active area of reservoir triggered earthquakes in India. *Geophysical Research Letters*, **30**(7), 1416, doi:10.1029/2002GLO016694.

Chadha, R. K., Srivastava, K., and Kumpel, H. J., 2005. Earthquake related changes in well water level and their relation to a static deformation model for the seismically active Koyna–Warna region, India. In Rummel, F. (ed.), *Rock Mechanics with Emphasis on Stress*. New Delhi: Oxford and IBH, pp. 135–150.

Chadha, R. K., Chandrani Singh, and Shekar, M., 2008. Transient changes in well water level in bore wells in western India due to 2004 Mw 9.3 Sumatra earthquake. *Bulletin of the Seismological Society of America*, **98**(5), 2553–2558, doi:10.1785/0120080952.

Chia, Y., Wang, Y. S., Chiu, J. J., and Liu, C. W., 2001. Changes of groundwater level due to the 1999 Chi-Chi earthquake in the Choshui River alluvial fan in Taiwan. *Bulletin of the Seismological Society of America*, **91**, 1062–1068.

Cooper, H. H., Bredehoeft, J. D., Papdopulos, I. S., and Bennett, R. R., 1965a. The response of well-aquifer systems to seismic waves. *Journal of Geophysical Research*, **70**, 3915–3926.

Cooper, H. H., Bredehoeft, J. D., Papdopulos, I. S., and Bennett, R. R., 1965b. The response of well-aquifer systems to seismic waves. *Journal of Geophysical Research*, **70**, 3915–3926.

Cooper, H., Bredehoeft, J. D., and Papadopulos, I. S., 1967. Response of a finite-diameter well to an instantaneous charge of water. *Water Resources Research*, **3**, 263–269.

Gavrinlenko, P., Melikadze, G., Chelidze, T., Gibert, D., and Kumsiashvili, G., 2000. Permanent water level drop associated with Spitak earthquake: observations at Lisi Borehole (Republic of Georgia) and modeling. *Geophysical Journal International*, **143**, 83–98.

Grecksch, G., Roth, F., and Kuempel, H. J., 1999. Coseismic well level changes due to the 1992 Roermond earthquake compared to static deformation of half space solutions. *Geophysical Journal International*, **138**, 470–478.

Gupta, H. K., Radhakrishna, I., Chadha, R. K., Kuempel, H.-J., and Grecksch, G., 2000. Pore pressure studies initiated in area of reservoir-induced earthquakes in India. *EOS. Transactions of the American Geophysical Union*, **81**(14), 145–151.

Harp, E. L., Jibson, R. W., Kayen, R. E., and Keefer, D. K., 2003. Landslides and liquefaction triggered by the Mw 7.9 Denali fault earthquake of 3 November 2002. *GSA Today*, **13**, 4–10.

Huang, W., Rojstaczer, S., and Breau, S., 1995. Coseismic response of water level to earthquakes in the San jacinto Fault, Southern California. *EOS, Transaction of the American Geophysical Union*, **76**(Suppl.), 359.

Huang, F.-Q., Jian, C.-L., Tang, Y., Xu, G.-M., Deng, Z.-H., and Chi, G.-C., 2004. Response changes of some wells in the mainland subsurface fluid monitoring network of China, due to the September 21, 1999, Ms 7.6 Chi-Chi earthquake. *Tectonophysics*, **390**(1–4), 217–234.

Igarashi, G., Wakita, H., and Sato, T., 1992. Precursory and coseismic anomalies in well water levels observed for the February 2, 1992 Tokyo Bay earthquake. *Geophysical Research Letters*, **19**, 1583–1586.

Janssen, J., 1998. Untersuchung des zusammenhangs zwishen lokaler seismizitat und dem porendruck in gespannten aquiferen in der Koyna region, Indien, *Diploma Thesis*, Germany: University of Bonn, 120 pp.

Kayen, R. E., Thompson, E., Minasian, D., Moss, R. E., Collins, B., Sitar, N., Dreger, D., and Carver, G., 2004. Geotechnical reconnaissance of 2002 Denali fault, Alaska earthquake. *Earthquake Spectra*, **20**, 639–667.

King, C.-Y., Azuma, S., Igarashi, G., Ohno, M., Saito, H., and Wakita, H., 1999. Earthquake-related water-level changes at 16 closely clustered wells in Tono, central Japan. *Journal of Geophysical Research*, **104**, 13073–13082.

King, C.-Y., Zuma, S., Hno, M., Asai, Y., Kitagawa, Y., Igarashi, G., and Wakita, H., 2000. In search of earthquake precursors in the water-level data of 16 closely clustered wells at Tono, Japan. *Geophysical Journal International*, **143**, 469–477.

Kissin, I. G., Belikov, V. M., and Ishankuliev, G. A., 1996. Short-term groundwater level variations in a seismic region as an indicator of the geodynamic regime. *Tectonophysics*, **265**, 313–326.

Kitagawa, Y., Koizumi, N., Takahashi, M., Matsumoto, N., and Sato, T., 2006. Changes in groundwater levels or pressures associated with the 2004 earthquake off the west coast of northern Sumatra (M 9.0). *Earth Planets Space*, **58**, 173–179.

Koizumi, N., Kano, Y., Kitagawa, Y., Sato, T., Takahashi, M., Nishimura, S., and Nishida, R., 1996. Groundwater anomalies associated with the 1995 Hyogoken Nanbu earthquake. *Journal of Physics of the Earth*, **44**, 373–380.

Koizumi, N., Tsukuda, E., Kamigaichi, O., Matsumoto, N., Takahashi, M., and Sato, T., 1999. Preseismic changes in groundwater level and volumetric strain associated with earthquake swarms off the east coast of the Izu peninsula, Japan. *Geophysical Research Letters*, **26**, 3509–3512.

Kuempel, H.-J., 1992. About the potential of wells to reflect stress variations within inhomogeneous crust. *Tectonophysics*, **211**, 317–336.

Kunugi, T., Fukao, Y., and Ohno, M., 2000. Underdamped responses of a well to nearby swarm earthquakes off the coast of Ito City, central Japan, 1995. *Journal of Geophysical Research*, **105**, 7805–7818.

Lai, W.-C., Koizumi, N., Matsumoto, N., Kitagawa, Y., Lin, C.-W., Shieh, C.-L., and Lee, Y.-P., 2004. Effects of seismic ground motion and geological setting on the coseismic groundwater level changes caused by the 1999 Chi-Chi earthquake Taiwan. *Earth Planets Space*, **56**, 873–880.

Liu, L.-B., Roeloffs, E., and Zheng, X.-Y., 1989. Seismically induced water level fluctuations in the WaliWell, Beijing, China. *Journal of Geophysical Research*, **94**, 9453–9462.

Liu, C., Huang, M. W., and Tsai, Y. B., 2006. Water level fluctuations induced by ground motions of local and teleseismic earthquakes at two wells in Hualien, eastern Taiwan. *TAO*, **17**, 371–389.

Matsumoto, N., Kitagawa, G., and Roeloffs, E. A., 2003. Hydrological response to earthquakes in the Haibara well, central Japan – I. Ground water level changes revealed using state space decomposition of atmospheric pressure, rainfall and tidal responses. *Geophysical Journal International*, **155**, 885–898.

Monakhov, F. I., Khantayev, A. M., Kechaykina, V. S., and Shishkin, A. A., 1983. Comparative characteristics of the response of groundwater level and outflow to earthquake preparation. *Izvestiya, Earth Physics*, **19**, 405–408.

Montgomery, D. R., and Manga, M., 2003. Streamflow and water well response to earthquakes. *Science*, **300**, 2047–2049.

Muir-Wood, R., and King, G. C. P., 1993. Hydrological signatures of earthquake strain. *Journal of Geophysical Research*, **98**, 22035–22068.

Okada, Y., 1992. Internal deformation due to shear and tensile faults in a half space. *Bulletin of the Seismological Society of America*, **82**(2), 1018–1040.

Quilty, E. G., and Roeloffs, E. A., 1997. Water level changes in response to the December 20, 1994 M 4.7 earthquake near Parkfield, California. *Bulletin of the Seismological Society of America*, **87**, 310–317.

Rexin, E. E., Oliver, J., and Prentiss, D., 1962. Seismically induced fluctuations of the water level in the Nunn-Bush well in Milwaukee. *Bulletin of the Seismological Society of America*, **52**, 17–25.

Roeloffs, E. A., 1988. Hydrologic precursors to earthquakes: a review. *Pure and Applied Geophysics*, **126**, 177–209.

Roeloffs, E. A., 1996. Poroelastic techniques in the study of earthquake related hydrological phenomena. *Advances in Geophysics*, **37**, 135–195.

Roeloffs, E. A., 1998. Persistent water level changes in a well near Parkfield, California due to local and distant earthquakes. *Journal of Geophysical Research*, **103**, 869–889.

Roeloffs, E. A., Sneed, M., Galloway, D. L., Sorev, M. L., Farrar, C. D., Howle, J. F., and Hughes, J., 2003. Water-level changes induced by local and distant earthquakes at Long Valley caldera. California. *Journal of Volcanology and Geothermal Research*, **127**, 269–303.

Rojstaczer, S., 1988a. Intermediate period response of water levels in wells to crustal strain: sensitivity and noise level. *Journal of Geophysical Research*, **93**, 13619–13634.

Rojstaczer, S., 1988b. Determination of fluid flow properties from response of water levels in wells to atmospheric loading. *Water Resources Research*, **24**, 1927–1938.

Rojstaczer, S., and Agnew, D. C., 1989. The influence of formation material properties on the response of water levels in wells to earth tides and atmospheric loading. *Journal of Geophysical Research*, **94**, 12403–12411.

Rudnicki, J. W., Yin, J., and Roeloffs, E. A., 1993. Analysis of water level changes induced by fault creep at Parkfield, California. *Journal of Geophysical Research*, **98**, 8143–8152.

Sadovsky, M. A., Monakhov, F. I., Kissin, I. G., and Shirokov, B. D., 1979. Short-term hydrogeodynamic precursors of

earthquakes in earthquake prediction. *Proceedings of the International (UNESCO) Symposium on Earthquake Prediction*. Paris, France.
Sil, S., 2006a. Response of Alaskan wells to near and distant large earthquakes. Master's Thesis, University of Alaska, 83 pp.
Sil, S., and Freymueller, J. T., 2006b. Well water level changes in Fairbanks, Alaska due to the great Sumatra–Andaman earthquake. *Earth Planets Space*, **58**, 181–184.
Singh, C., Ramana, D. V., Chadha, R. K., and Shekar, M., 2008. Coseismic responses and the mechanism behind Mw 5.1 earthquake of March 14, 2005 in the Koyna–Warna region, India. *Journal of Asian Earth Sciences*, **31**(4–6), 499–503.
Sterling, A., and Smets, E., 1971. Study of earth tides, earthquakes and terrestrial spectroscopy by analysis of the level fluctuations in a borehole at Heibaart (Belgium). *Geophysical Journal of the Royal Astronomical Society*, **23**, 225–242.
Stricherz, V., 2003. Seismology getting to know hydrology. *University Week*, **29**, 20.
Talwani, P., 1997. Seismotectonics of the Koyna–Warna area, India. *Pure and Applied Geophysics*, **150**, 511–550.
Van der Kamp, G., and Gale, J. E., 1983. Theory of the earth tide and barometric effects in porous formations with compressible grains. *Water Resources Research*, **19**, 538–544.
Wakita, H., 1975. Water wells as possible indicators of tectonic strains. *Science*, **189**, 553–555.
Wang, C., and Manga, M., 2006. Threshold distance for streamflow increase and liquefaction during earthquakes – implications on their mechanisms. *Western Pacific Geophysics Meeting Supplement*, Vol. 87, No. 36, Beijing, China, 24–27 July 2006.
Wang, C. Y., Cheng, L. H., Chin, C. V., and Yu, S. B., 2001. Coseismic hydrologic response of an alluvial fan to the 1999 Chi-Chi earthquake, Taiwan. *Geology*, **29**, 831–834.
Wesson, R. L., 1981. Interpretation of changes in water level accompanying fault creep and implications for earthquake prediction. *Journal of Geophysical Research*, **86**, 9259–9267.
Zhang, Z.-D., Chi, Z. L., Chen, H. M., and Luan, G.-Y., 2000. Vibration of water level in well and earthquake wave. *Journal of Seismological Research*, **23**(4), 418–425.

Cross-references

Earthquake Precursors and Prediction
Poroelasticity
Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity

SEISMIC STRUCTURE AT MID-OCEAN RIDGES

Donald W. Forsyth
Department of Geological Sciences, Brown University, Providence, RI, USA

Definition

Lithosphere. The cool, uppermost layer of the solid earth that moves as a unit and has some long-term elastic rigidity. It usually consists of both crust and the uppermost mantle. The mantle portion is typically characterized by high seismic velocities and low attenuation.
Asthenosphere. A more deformable, low-viscosity layer in the mantle underlying the lithosphere. Characterized by low seismic velocities and relatively high attenuation.
Anisotropy. Physical properties at any one point varying depending on direction; for seismic waves, velocity may depend on direction of propagation and direction of polarization of the wave.

Introduction

Mid-ocean ridges are spreading centers where two oceanic, lithospheric plates move apart. The separation of the plates induces upwelling in the underlying asthenospheric mantle. Because melting temperature decreases with decreasing pressure, as the mantle upwells, it undergoes pressure-release partial melting, producing magma that migrates upward to form new oceanic crust. As the plates move away from the spreading center, heat is lost to the surface conductively by diffusion and convectively by hydrothermal circulation. As the crust and mantle lose heat, the magma solidifies, and the plates thicken and cool with increasing age of the seafloor.

Although the general outline of the processes beneath mid-ocean ridges leading to the formation of new seafloor is well known, there are many questions remaining about the details of the mantle flow, melt generation, and melt migration that have not yet been fully answered. For example, along mid-ocean ridges, spreading centers are offset and segmented by transform faults and overlapping spreading centers. Are there distinct centers of mantle upwelling beneath each ridge segment or is the segmentation a shallow manifestation of stresses within the lithosphere with upwelling a more or less continuous phenomenon in the asthenosphere? How broad and deep is the melt production region?

The propagation of seismic waves through the crust and mantle provides one of the most direct ways of probing the structure beneath the ridges associated with plate separation and crust formation. The velocity, attenuation, and anisotropy of the waves are affected by temperature, composition, the presence of melt or cracks, and the crystal fabric. Tomographic images and maps of crustal and uppermost mantle structure are typically created in active source experiments where artificial sound sources, such as air guns, are recorded by ocean bottom seismometers (OBS) and/or long arrays of hydrophones towed behind ships. Deeper mantle structure is probed using signals generated by distant, teleseismic earthquakes recorded on arrays of ocean-bottom and land seismometers over periods of months to years. The logistical difficulty and expense of such experiments means that relatively few mid-ocean ridges have been studied in detail.

Mantle structure

A cross section of the shear velocity structure beneath the East Pacific Rise spreading center is shown in Figure 1 in comparison to the velocity variations that would be expected for simple conductive cooling of the plates if shear velocity were sensitive only to temperature and pressure. This tomographic image is based on the propagation of Rayleigh surface waves across two OBS arrays deployed for periods of

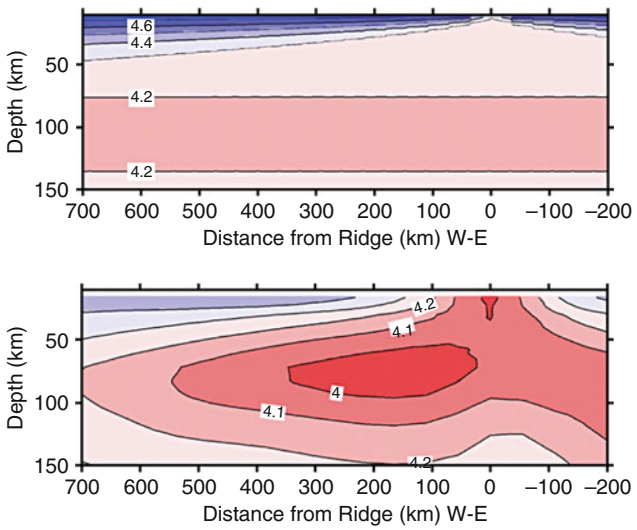
6 and 12 months. The expected thickening of the plate with increasing distance from the ridge axis as the plate cools is clearly observed in the form of increasing shear wave velocity near the surface. Velocity changes extend substantially deeper than is predicted and the shear velocity is lower than expected for the direct effects of temperature variations. Both of these departures from the predictions are indications that partial melt may be present, which could reduce the shear (S) wave velocity.

The lowest velocities are observed at depths of 60–80 km, where petrological models predict maximum melt production. There also are very low shear velocities at shallow depths immediately beneath the ridge axis, which may represent higher melt concentrations in the mantle that accumulated as melt migrates upward and inward to form the new oceanic crust at the ridge axis.

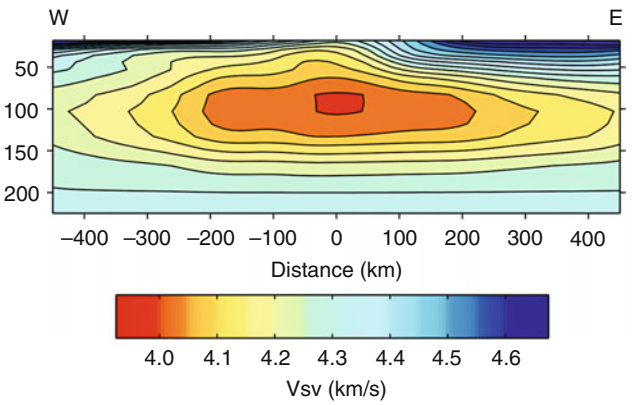
Another important feature of the shear velocity structure is the asymmetry across the ridge axis. To the east, beneath the Nazca plate, the high-velocity surface layer increases in thickness more rapidly and the very low-velocity region is absent. This asymmetry is also observed in the degree of shear wave splitting, an indicator of anisotropy, in delays of compressional (P) and S waves, in electrical conductivity, and in the rate of subsidence of the seafloor away from the ridge axis (MELT Seismic Team, 1998; Evans et al., 1999). S wave tomography indicates that the asymmetry may extend to depths of 200 km or more (Hammond and Toomey, 2003). The asymmetry is probably caused by large-scale mantle flow coming from the hotspot region to the west beneath the Pacific plate, coupled with migration of the spreading center to the west. The fast direction for seismic wave propagation, as indicated by shear wave splitting and Rayleigh wave anisotropy, is perpendicular to the East Pacific Rise, consistent with the alignment of olivine crystals in the mantle expected for plate formation and flow from the west.

The East Pacific Rise is one of the fastest spreading ridges, with full spreading rate of about 14 cm/year. Pronounced asymmetry is also observed across the Reykjanes Ridge south of Iceland (Figure 2), which has a full spreading rate of only about 2 cm/year (Delorey et al., 2007). In this case, the tomographic study took advantage of the existence of arrays of stations on Iceland, which straddles the Reykjanes/Mid-Atlantic Ridge. Like the East Pacific Rise, very low shear velocities (~4.0 km/s) indicative of the presence of melt are found in a broad region beneath the ridge. For both the East Pacific Rise and the Reykjanes Ridge, there is too little attenuation of surface waves to attribute the very low velocities to the effect of high temperature alone. Velocities at depths shallower than 80 km are lower on the west side, beneath the North American plate, perhaps due to the westward migration of the ridge and upwelling in the mantle in the wake of the thicker North American lithosphere farther west. The anisotropy pattern is different than for typical mid-ocean ridges, perhaps indicating that there is along-axis flow away from the Iceland hotspot in the asthenosphere.

To date, there have been no experiments that provide good control of along-axis variations in mantle structure at depths of tens of kilometers or more where melt production is expected to occur. In the Gulf of California, a surface wave study showed that there are along-axis variations in shear velocity with minima spaced at intervals of about 250 km, perhaps indicative of discrete upwelling centers (Wang et al., 2009), but that spreading system is flanked by continental crust on both sides and dominated by long transform faults, so it may not be typical. At shallower depths just



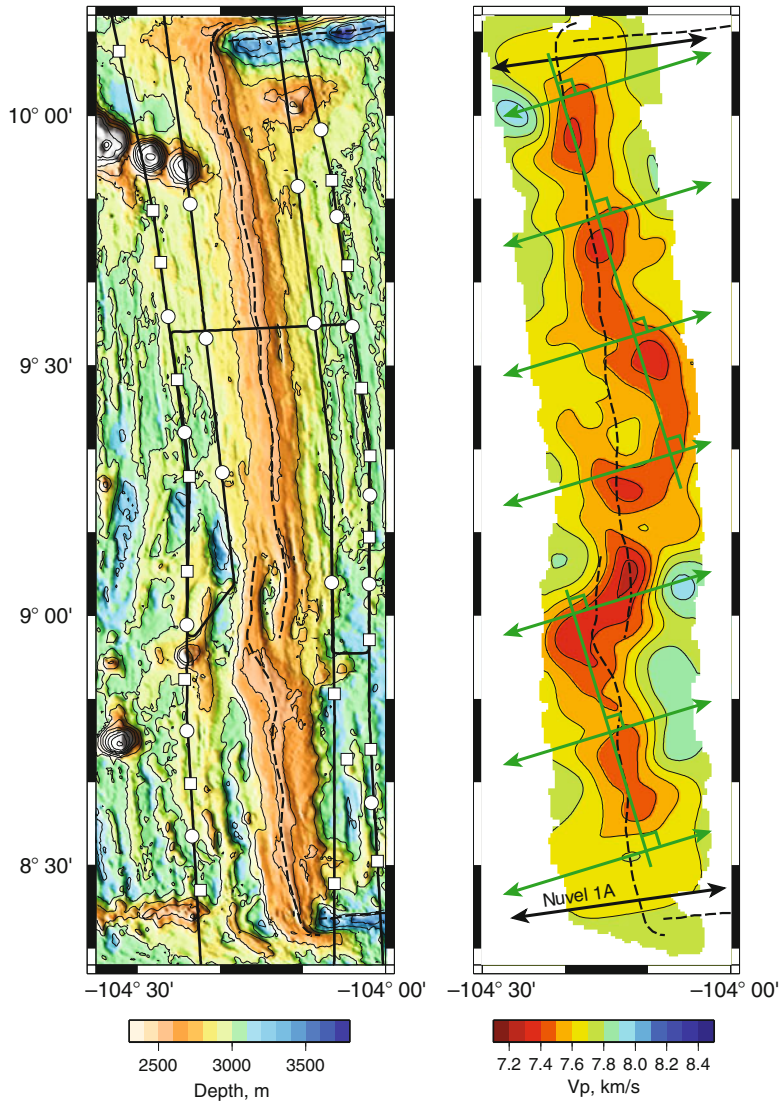
Seismic Structure at Mid-Ocean Ridges, Figure 1 Tomographic cross section of the East Pacific Rise, comparing predicted shear velocity structure (*top*) to observed structure (*bottom*). Contours are labeled in km/s. Velocities are significantly lower than predicted for temperature effects alone, suggesting the presence of partial melt. After Harmon et al. (2009).



Seismic Structure at Mid-Ocean Ridges, Figure 2 Tomographic cross section of shear wave velocity structure across Reykjanes Ridge south of Iceland. Note strong asymmetry in upper 100 km between North American plate (west) and European plate (east). After Delorey et al. (2007).

beneath the Moho (the transition from crust to mantle), there are clearly along-axis variations in P-wave velocity on the northern East Pacific Rise, the boundary between the Pacific and Cocos plates (Toomey et al., 2007). Using long offset arrivals refracting from the Moho (Pn) observed in an active experiment, velocity minima were found spaced about 25 km apart (Figure 3). P-wave velocities in the minima are 7.4 km/s or less, suggesting the presence of 1–3% melt distributed in films or thin sheets (typical P-wave velocities at the Moho are 7.8–8.4 km/s). However, it is not clear whether these apparent centers of melt lie above centers of upwelling mantle or they represent a scale length associated with melt migration.

Most of the centers of melt concentration lie beneath or very close to the spreading center determined from detailed bathymetric surveys, but one at 9° 30' N is displaced several kilometers from the axis. Because the fast direction for P-wave propagation is skewed from the spreading direction and is not perpendicular to the strike of the ridge axis, Toomey et al., inferred that upwelling and mantle flow at depth may also be skewed, controlling the location of the velocity minima just below the Moho.



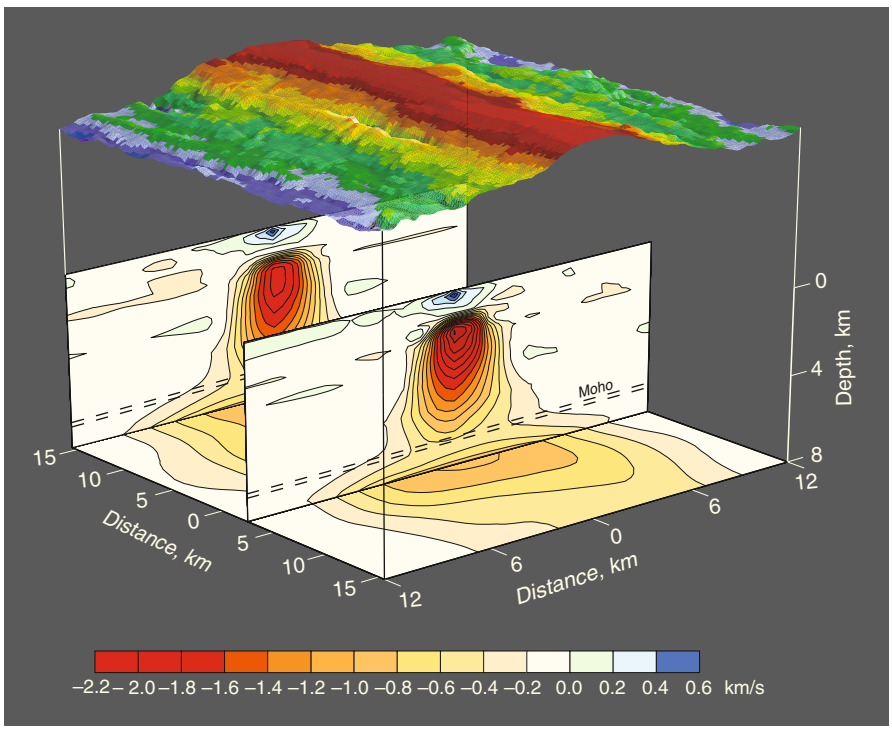
Seismic Structure at Mid-Ocean Ridges, Figure 3 Bathymetry of the East Pacific Rise (*left*) and tomographic image of the mantle 9 km beneath the seafloor (*right*). Dashed lines show plate boundary. Solid lines show locations of air gun shots fired at 500-m intervals. Squares and circles are locations of ocean bottom receivers. Contour interval on tomographic image of P-wave velocity is 0.1 km/s. Green lines with double arrowheads indicate fast direction for anisotropic wave propagation. Black lines with arrows indicate direction of relative motion between the Pacific and Cocos plates. Note that some of the slowest region is displaced from the plate boundary. After Toomey et al. (2007).

Crustal structure

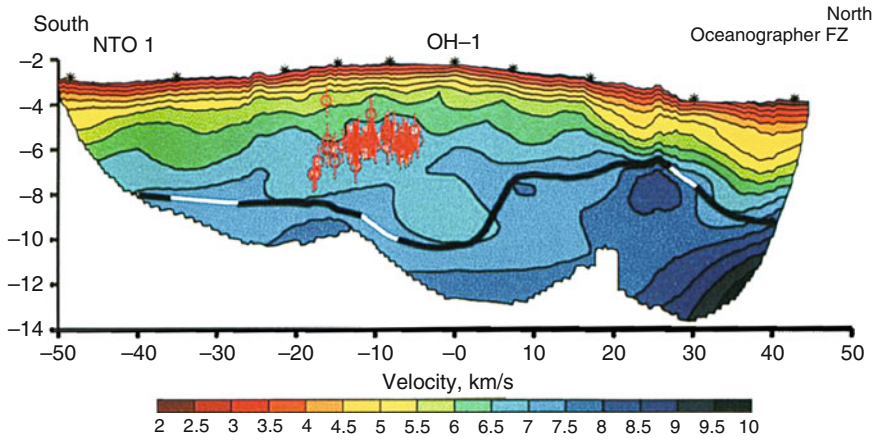
Basaltic melt migrates upward through the mantle and is focused at the ridge axis. The mechanisms for focusing are still not well known: there may be melt-rich channels at the base of the lithosphere that guide the melt upward and toward the ridge axis; there may be pressure gradients within the deforming mantle that help push the melt toward the axis; or there may be anisotropic cracks or dunite channels that form easy paths for melt migration. Once the melt reaches crustal levels, there may be redistribution along axis through dikes or a continuous magma chamber. It is clear from seismic studies, however, that there is very little magmatic addition to the crust outside the immediate vicinity of the ridge axis. The crust is essentially full thickness at the spreading center itself (Detrick et al., 1987).

The classic model of oceanic crust is a layer of extrusive basalts at the surface in the form of porous pillow basalts and sheet flows, underlain by a region of sheeted dikes that feed the extrusive layer from a magma chamber, and the lower crust consisting of gabbros that solidify from the magma chamber or underlying mush zone. The seismic structure has also commonly been described in terms of layers that have often been equated to the lithological layering: layer 2A is a low-velocity layer at the surface several hundred meters thick (layer 1 is sediments that are deposited on top of the basaltic crust); layer 2B is a transition region in which the P velocity increases rapidly downward; and layer 3 has low vertical velocity gradients and high P-wave velocities of 6.5–7.0 km/s. A number of investigations, however, have demonstrated that the seismologic layering does not correspond exactly to the lithological layering and that seismic structure is primarily controlled by porosity. The middle to lower crust, layer 3, contains both dikes and gabbro bodies.

At the axis of fast-spreading ridges, there usually is a low-velocity region only a few kilometers wide at most that represents a zone of partial melting that extends throughout the crust (Figure 4). At the top of this low-velocity region is a very low velocity layer, which is imaged as a prominent reflector in seismic reflection profiles at depths of 1.5–2.5 km on the East Pacific Rise (Detrick et al., 1987). The depth to this reflector tends to increase with decreasing spreading rate or proximity to a fracture zone and it is typically absent at slow-spreading ridges (Singh et al., 2006). The width varies from about 250 m to several kilometers. It is interpreted as the top of an axial magma chamber (AMC) or melt sill, with typical thickness less than 100 m. Modeling of P to S conversions in wide-angle reflections indicates that melt content in the AMC varies from nearly 100% to less than 30% along the ridge (Singh et al., 1998; Canales et al., 2006). Estimates of melt content in the deeper crustal mush zone (consisting



Seismic Structure at Mid-Ocean Ridges, Figure 4 Three-dimensional view of the anomalous P-wave velocity structure of the East Pacific Rise at 9°30′N (a one-dimensional reference model has been subtracted). From Dunn et al. (2000). Overlying bathymetry is shown at top, with red shallow to blue, deep. The red regions below the ridge axis correspond to areas of partial melting. The axial magma chamber is at the top of these anomalously slow regions.



Seismic Structure at Mid-Ocean Ridges, Figure 5 P-wave velocity structure along the Mid-Atlantic Ridge axis from a non-transform offset on the left to the Oceanographer fracture zone on the right. Crustal thickness from wide-angle reflections off the Moho is shown by thick solid line (open where there are gaps in coverage). The crust is thicker and the lower crust has lower velocity near the center of the segment. From Hooft et al. (2000). The red circles are locations of microearthquakes with uncertainties indicated by error bars. Location of ocean-bottom hydrophones indicated by asterisks.

of a mixture of melt and crystals) are of the order of 2–8% in the lower crust and 3–12% near the Moho (Dunn et al., 2000). The fact that the zone in which melt is present is so narrow even at fast-spreading ridges means that hydrothermal circulation must be very efficient in removing heat, as purely conductive cooling of the crust would result in a much wider zone of partial melting.

Although there are variations in structure along-axis at fast-spreading ridges, the along-axis variations are much more pronounced at segmented, slow-spreading ridges. Typically, the crust thins approaching transform offsets. Within the fracture zone itself, the basaltic crust may be as thin as 1 km or less, but seismically there may be an altered, fractured layer that is a few kilometers thick and characterized by unusually low velocities, so that it looks like crust. The low-velocity region in the fracture zone, however, probably is mostly mantle that is altered by interaction with water penetrating down cracks that are repeatedly opened by slip along the transform fault. Near the center of ridge segments between two transform offsets, the crust tends to be thicker and lower in velocity than elsewhere (Figure 5), suggesting that melt is preferentially delivered to the crust from the mantle at that point. The upper crust at slow spreading ridges is anisotropic, with P-waves traveling faster along axis than perpendicular to it, indicating that faults and fissures are preferentially aligned parallel to the spreading center (Barclay and Toomey, 2003).

Summary

The seismic velocity structure of mid-ocean ridges is controlled by crustal thickness, cracking or porosity, temperature, melt, and crystal orientation. Low-velocity regions in the mantle that are caused by high temperatures and partial melt indicate that melt is generated in a broad region beneath spreading centers. Asymmetries in the velocity structure show that upwelling and melting beneath ridges are strongly influenced by global mantle circulation and plate motions. Crustal low-velocity regions are concentrated very near the ridge axis, so melt must migrate both vertically and horizontally to the ridge axis from the broad melt production region. The narrowness of the low-velocity zone in the crust requires that hydrothermal circulation must rapidly cool the crust. Along-axis variations, particularly at slow-spreading ridges, suggest that melt is preferentially delivered to the center of ridge segments.

Bibliography

Barclay, A. H., and Toomey, D. R., 2003. Shear wave splitting and crustal anisotropy at the Mid-Atlantic Ridge, 35°N. *Journal of Geophysical Research*, **108**, doi:10.1029/2001JB000918.

Canales, J. P., Singh, S. C., Detrick, R. S., Carbotte, S. M., Harding, A., Kent, G. M., Diebold, J. B., Babcock, J., and Nedimovic, M. R., 2006. Seismic evidence for variations in axial magma chamber properties along the southern Juan de Fuca Ridge. *Earth and Planetary Science Letters*, **246**, 353–366.

Delorey, A. A., Dunn, R. A., and Gaherty, J. B., 2007. Surface wave tomography of the upper mantle beneath the Reykjanes ridge with implications for ridge-hot spot interaction. *Journal of Geophysical Research*, **112**, B08313, doi:10.1029/2006JB004785.

Detrick, R. S., Buhl, P., and Vera, E., 1987. Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature*, **326**, 35–41.

Dunn, R. A., Toomey, D. R., and Solomon, S. C., 2000. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the east pacific rise at 9 degrees 30'N. *Journal of Geophysical Research*, **105**, 23537–23555.

Evans, R. L., et al., 1999. Asymmetric electrical structure in the mantle beneath the east pacific rise at 17°S. *Science*, **286**, 756–759.

Hammond, W. C., and Toomey, D. R., 2003. Seismic velocity anisotropy and heterogeneity beneath the mantle electromagnetic and tomography experiment (MELT) region of the East Pacific Rise from analysis of P and S body waves. *Journal of Geophysical Research*, **108**, doi:10.1029/2002JB001789.

Harmon, N., Forsyth, D. W., and Weeraratne, D. S., 2009. Thickening of young Pacific lithosphere from high-resolution Rayleigh

wave tomography: a test of the conductive cooling model. *Earth and Planetary Science Letters*, **278**, 96–106.
Hooft, E. E. E., Detrick, R. S., Toomey, D. R., Collins, J. A., and Lin, J., 2000. Crustal thickness and structure along three contrasting spreading segments of the Mid-Atlantic Ridge, 33.5° – 35°N. *Journal of Geophysical Research*, **105**, 8205–8226.
MELT Seismic Team, 1998. Imaging the deep seismic structure beneath a Mid-Ocean Ridge: the MELT experiment. *Science*, **280**, 1215–1218.
Singh, S. C., Crawford, W. C., Carton, H., Seher, T., Combier, V., Cannat, M., Canales, J. P., Dusunur, D., Escartin, J., and Miranda, J. M., 2006. Discovery of a magma chamber and faults beneath a Mid-Atlantic Ridge hydrothermal field. *Nature*, **442**, 1029–1032.
Singh, S. C., Kent, G. M., Collier, J. S., Harding, A. J., and Orcutt, J. A., 1998. Melt to mush variations in crustal magma properties along the ridge crest at the southern East Pacific Rise. *Nature*, **394**, 874–878.
Toomey, D. R., Jousselin, D., Dunn, R. A., Wilcock, W. S. D., and Detrick, R. S., 2007. Skew of mantle upwelling beneath the East Pacific Rise governs segmentation. *Nature*, **446**, 409–414.
Wang, Y., Forsyth, D. W., and Savage, B., 2009. Convective upwelling in the mantle beneath the Gulf of California. *Nature*, **462**, 499–501, doi:10.1038/nature08552.

Cross-references

Crustal Reflectivity (Oceanic) and Magma Chamber
Earth's Structure, Upper Mantle
Lithosphere, Mechanical Properties
Lithosphere, Oceanic
Lithosphere, Oceanic: Thermal Structure
Ocean Bottom Seismics
Seafloor Spreading
Seismic Anisotropy
Seismic Data Acquisition and Processing
Seismic Tomography
Seismic Velocity-Density Relationships
Surface Waves
Traveltime Tomography Using Controlled-Source Seismic Data

SEISMIC TOMOGRAPHY

Guust Nolet
Geosciences Azur, Sophia Antipolis, France

Definition

The term tomography derives from the Greek τομος, or slice. "Seismic tomography" is used for a variety of methods that use transmitted seismic waves to estimate the spatial variations in properties (wave velocity, density, attenuation) inside the Earth, which are often represented as images of two-dimensional cross-sections or "slices." It is conceptually different from seismic migration, which uses reflected waves to image sharp discontinuities.

History

In 1971, P. Bois at the Institut Français de Pétrole was the first to suggest the tomographic method in order to locate the causes of delays in seismic waves between two boreholes. His paper predates many future developments but was written in French and remained largely unnoticed. In the mid-1970s, Keiti Aki from MIT applied a linear inversion to locate velocity heterogeneities beneath large nuclear monitoring arrays in Norway and Montana, and Harvard's Adam Dziewonski began interpreting the time residuals published by the International Seismological Center (ISC) in the UK in terms of global velocity anomalies.

In 1982, Guy Masters and his colleagues at the Scripps Institution of Oceanography discovered a strong degree-2 component in the geographical distribution of the slight shifts in the spectral peaks of the Earth's normal modes. Since then, the Earth's free oscillations have contributed to constrain the heterogeneity in the Earth at the longest wavelengths and as deep as the inner core. By 1984 John Woodhouse and Adam Dziewonski at Harvard published a first global model for shear velocity in the upper mantle based on long-period surface waves.

However, to image smaller scale anomalies, the shorter wavelengths of P and S-waves are indispensable. In particular, Steve Grand at the University of Texas, Rob van der Hilst and Wim Spakman at Utrecht University and Yoshio Fukao and colleagues at the University of Tokyo pioneered high-resolution body-wave tomography using iterative solvers for the huge systems of linearized equations and established in the early 1990s that some, but not all, subducting slabs are able to sink well into the depths of the lower mantle.

Thermal plumes in the lower mantle were for the first time reliably imaged in 2003 by Raffaella Montelli, using a new technique of finite-frequency tomography developed by Tony Dahlen and Guust Nolet and their collaborators at Princeton University.

For references and a detailed account of the history of seismic tomography see the reviews by Romanowicz (2003) and Rawlinson et al. (2010). Nolet (2008) provides a general introduction into the methods of seismic tomography, including the theoretical aspects that are here discussed only briefly.

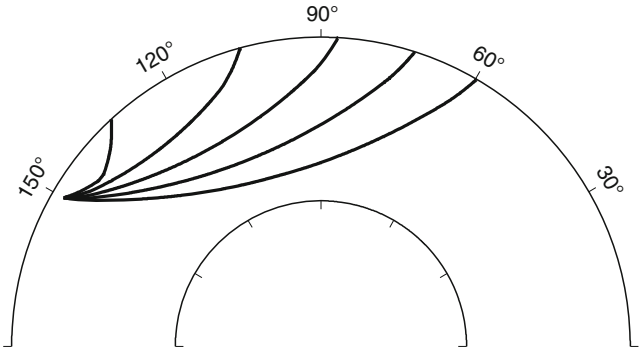
Onset times

Much of seismic tomography is based on estimating the arrival time of a seismic body wave by picking the "onset" of a phase on the seismogram, and interpreting the travel time T using the infinite-frequency approximation of ray theory:

$$T = \int_{\text{raypath}} \sigma(\boldsymbol{r}) \mathrm{d}s \tag{1}$$

where the raypath is determined by Snell's law and where σ is the inverse velocity v^{-1} of the wave at the location $\boldsymbol{r}$. The raypaths may be located between two boreholes, e.g., to monitor the exploitation of an oil or gas field, between several explosive sources and an array of seismographs at the surface (deep seismic sounding), between a seismogenic zone and a local array of seismographs, or between an earthquake and the global network of seismic stations.

The raypath is often approximated by the path calculated for a spherically symmetric Earth or a horizontally



Seismic Tomography, Figure 1 Examples of raypaths for a P wave in an equatorial cross-section from a hypothetical earthquake at a depth of 248 km located at 0°N, 150°E.

layered background model. Examples of global raypaths are given in Figure 1. This approximation is permitted because ray trajectories render the travel time stationary, such that small deviations in the true ray path location only cause second order errors in the travel time calculated with Equation 1, a property of rays known by the name of *Fermat's Principle*. In practice, one usually prefers to invert for the difference between the true Earth and a starting- or background model: $\delta\sigma(\boldsymbol{r}) = \sigma(\boldsymbol{r}) - \sigma_0(\boldsymbol{r})$, and use the difference between observed time T and its prediction T_0 from the background model as datum: $\delta T = T - T_0$:

$$\delta T = \int_{\text{raypath}} \delta\sigma(\boldsymbol{r})\mathrm{d}s \tag{2}$$

For strongly heterogeneous regions such as subduction zones, one may need to use three-dimensional ray tracing and an iterative approach using (Equation 2).

Model parameterization and inversion

The model $\delta\sigma(\boldsymbol{r})$ can be described by a finite number of parameters if it is developed in terms of a basis of M interpolating or "basis" functions $h_k(\boldsymbol{r})$:

$$\delta\sigma(r) = \sum_{k=1}^{M} m_k h_k(\boldsymbol{r}) \tag{3}$$

The basis functions may be represented by homogeneous cells, linear interpolators in a tetrahedral or other mesh, spherical harmonics, or 3D wavelets. Substitution of Equation 3 in Equation 2 gives a formal system of linearized equations for N estimated travel times $\delta T_i, i = 1, \ldots N$, arranged in a vector $\delta\boldsymbol{T}$:

$$\delta\boldsymbol{T} = \boldsymbol{A}\boldsymbol{m} \tag{4}$$

where $A_{ik} = \int_{\text{raypath}_i} h_k(\boldsymbol{r})\mathrm{d}\boldsymbol{s}$. Since many raypaths may overlap, the system (Equation 4) is usually overdetermined and needs to be solved by a least squares solver that minimizes χ^2, the length of the misfit vector weighted by the data standard error e_i:

$$\chi^2(\boldsymbol{m}) = \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{M} A_{ik} m_k - \delta T_i}{e_i} \right)^2 \tag{5}$$

Mathematically, this is accomplished by first dividing the system (Equation 4) by the standard error of the data (i.e., multiplying by the inverse square root of the covariance matrix $\boldsymbol{C}$ which is generally assumed to be diagonal), and backprojecting the system using the transpose $\boldsymbol{A}^T$ of $\boldsymbol{A}$:

$$\boldsymbol{A}^T\boldsymbol{C}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{m} = \boldsymbol{A}^T\boldsymbol{C}^{-\frac{1}{2}}\boldsymbol{\delta T} \tag{6}$$

One often scales the equations a priori such that the data have unit error $e_i = 1$ and $\boldsymbol{C} = \boldsymbol{I}$, which has the same effect. In the following we therefore ignore $\boldsymbol{C}$.

Usually, Equation 6 is at the same time overdetermined for some parameters and underdetermined for others (the determinant of $\boldsymbol{A}^T\boldsymbol{A}$ being zero): there are in that case infinitely many models that satisfy the data within the measurement uncertainty and one needs to regularize the solution. This can be done by choosing the solution that minimizes a weighted penalty of χ^2, model norm and model roughness, by strongly truncating a development in spherical harmonics, or by choosing the sparsest decomposition in wavelets that still satisfies the data. This generally involves a subjective choice between the degree of detail one allows in a model and the goodness of fit to the data. Generally, one aims for a χ^2 approximately equal to the number of data N – in other words, one attempts to fit the data at the level of about one standard deviation. Invariably, a trade-off exists between the detail allowed in the model and the statistical precision with which the model parameters can be determined: the sharper the detail, the more uncertain the m_k.

Regularization can also be done in a fundamentally different way by inverting for local averages in $\delta\sigma(\boldsymbol{r})$ that can be estimated linearly from the data with a specified variance, though this can only be done at the expense of a significant increase in computing time (Backus-Gilbert theory). In this case the trade-off is between the size of the averaging volume and the model variance: averages over larger volumes are determined with smaller statistical uncertainty.

Whatever method is used, the system (Equation 4) may be very large (e.g., 10^6 data for 10^5 unknown model parameters). Local parameterizations (cells, as opposed to spherical harmonics) render $\boldsymbol{A}$ sparse, and the system can be solved efficiently using iterative solvers that adapt to sparse matrices, such as LSQR.

Normal modes and surface waves

The eigenfrequencies ${}_n\omega_\ell^m$ of the Earth are characterized by three quantum numbers ℓ, m and n, related to the number of nodal surfaces in the displacement field of the Earth with latitude, longitude and depth, respectively. For a non-rotating, isotropic, spherically symmetric Earth the spectrum is degenerate in the sense that the frequency is

independent of the azimuthal order m. For the real Earth, a weak dependence on m splits each eigenfrequency into $2\ell + 1$ separate frequencies that are too closely spaced to be resolvable except for the very lowest angular order ℓ. Instead, a composite spectral line or "multiplet" is observed with a peak that depends on the location of the seismic station – a direct consequence of the fact that the $2\ell + 1$ single peaks have amplitudes that depend on geographical location by virtue of their spherical harmonic dependence on latitude and longitude.

Two major strategies exist to exploit the small fluctuations in the spectrum. Decomposing the free oscillation into surface waves traveling in opposite directions, ray theory may be used to establish a linear relationship between the heterogeneity along the great circle between source and receiver and the observed peak shift in the spectral line. At higher frequency we separate single passages of the surface wave and exploit the linear relationship between fluctuations in the observed phase velocity and the Earth's heterogeneity. In both cases, the relationship between the Earth's heterogeneity and the observed datum is a two-dimensional integral along the great circle of the form:

$$\delta\omega = \int_0^a \int_{\mathrm{gc}} K(\boldsymbol{r})\delta\sigma(\boldsymbol{r})\mathrm{d}s\mathrm{d}r, \tag{7}$$

where the kernel $K(\boldsymbol{r})$ is computed using first order perturbation theory of the differential equations governing the Earth's free oscillations.

Alternatively, we may exploit the known distribution of amplitudes of single peaks over the surface of the Earth to invert for the location of these peaks ("mode splitting"). The small frequency shifts $\delta\omega^m$ are themselves the eigenvalues of a splitting matrix $\boldsymbol{H}$ of which the elements are linearly related to the variation of density and elastic parameters in the Earth. We can estimate $\boldsymbol{H}$ from the seismic data using autoregressive filtering techniques. This way we avoid any ray-theoretical approximations and obtain a three-dimensional integral constraint that can be used to solve for the large-scale variations in the Earth's density and elastic properties.

Finite-frequency tomography

Modern, broadband digital instrumentation allows for a robust estimation of the delay of an observed seismic wave $s_1(t)$ with respect to a theoretically predicted ("synthetic") waveform $s_2(t)$ or with respect to the same phase observed elsewhere, by locating the maximum in the cross-correlation $C(t)$ between the two signals:

$$\begin{aligned} \delta T &= \arg \max_{t\in[\tau_1,\tau_2]} C(t), \\ C(t) &= \int_{\tau_1}^{\tau_2} s_1(\tau)s_2(\tau - t)\,\mathrm{d}\tau\,, \end{aligned} \tag{8}$$

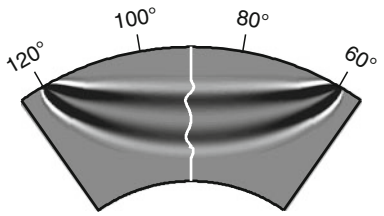
where the integration interval (τ_1, τ_2) extends over all times where the integrand is nonzero. The cross-correlation delay is thus fundamentally different from the delay measured by picking the onset of $s_1(t)$, because it represents an integral measure over a time window that is at least as long as the dominant period of $s_1(t)$. Heterogeneities inside the Earth may influence the integral by scattering waves in the direction of the recording station that arrive within the time window. The most serious consequence is that energy may diffract around a small heterogeneity, thus masking or dominating the slower or faster arrival that has crossed the heterogeneity. This phenomenon of "wavefront healing" biases observed delay times to zero, especially if the anomaly is located at some distance from the source and receiver.

Taking wave diffraction into account leads to a three-dimensional integral constraint of the form:

$$\delta T = \int_V K(\boldsymbol{r})\delta\sigma(\boldsymbol{r})\mathrm{d}^3\boldsymbol{r}\,, \tag{9}$$

where the kernel $K(\boldsymbol{r})$ can be efficiently calculated using ray theory for scattered waves. The effective volume of integration V is roughly equivalent to that of the Fresnel zone of the ray, where the sensitivity is largest. An example of a finite-frequency kernel is shown in Figure 2. Remarkably, the cross-correlation travel time is insensitive to perturbations in the Earth's properties at the location of the ray path. Because of this hole in the sensitivity and their general shape, finite-frequency kernels are often referred to as "banana-doughnut" kernels. Numerical tests have shown that the finite-frequency kernels accurately model the loss of signal caused by "wavefront healing," the gradual reduction of a delay caused by waves diffracting around a small heterogeneity. By estimating delays in different frequency bands, one gains information about the size of the heterogeneity in the Earth.

Though finite-frequency kernels were initially proposed to provide a better theoretical modeling of long-period signals affected by wavefront healing, they offer the advantage of being able to model the healing as a function of frequency, and thus obtain independent information on the size of the heterogeneity. Such "multiple-frequency" tomography may significantly increase



Seismic Tomography, Figure 2 The sensitivity of the cross-correlation time to the earth structure, for a long-period P wave from a surface source at 0°N, 120°E, recorded at a distance of 60°. The kernel is plotted in an equatorial cross-section of the mantle, the *grayscale* runs from large negative values (*black*) to positive (*white*). The sensitivity is zero at the location of the geometrical raypath, forming a "doughnut hole" in the banana-shaped kernel.

resolution (Sigloch et al., 2008). Finite-frequency sensitivity can also be formulated for the amplitude perturbations caused by attenuation and by focusing/defocusing (focusing cannot be handled with ray theory because ray theory does not model amplitudes correctly at low frequency and is very nonlinear at high frequency).

Instead of inverting for a cross-correlation delay or an amplitude perturbation, one can formulate a 3D sensitivity directly for the observed time series, though this has the disadvantage that the observed and predicted waveforms have to be close enough in phase that the phase difference can be adequately modeled by adding a small perturbation to the waveform itself (in fact modeling $e^{i\varphi} \approx 1 + i\varphi$). This limits the inversion to waveforms with small time mismatches, in contrast to the delays estimated through (Equation 8), which remain linear for large anomalies.

All types of data – delays, amplitudes and waveform mismatches – can be iteratively inverted using an adjoint approach, in which the best fitting model is sought along the gradient $\nabla_m \chi^2$ (Tromp et al., 2005), e.g., for delay times scaled to unit variance:

$$\nabla_m \chi^2 = \boldsymbol{A}^T(\boldsymbol{Am} - \delta \boldsymbol{T}) \tag{10}$$

For waveforms, the matrix A is in this case the matrix representation of a finite-difference or spectral-element algorithm that produces the predicted seismograms and A^T is the adjoint operator that projects the seismograms back in time. Since this allows one to backproject all residual seismograms for one earthquake with only one calculation, and since only the product of the adjoint matrix and the residual data vector is needed, this allows us to avoid computing the partial derivatives separately for each datum δT_i with respect to each model parameter m_k. The adjoint approach has some advantages over a direct matrix inversion for large-scale problems, in particular those with few sources and many stations, because it does not require the storage of a large matrix, which is especially important for waveform inversions.

Summary

The equations of seismic tomography are integral equations; the observations are weighted averages of the properties of the Earth (Equation 9). Though often treated as line integrals, assuming ray theory is valid, even a 1 Hz P wave senses a volume or Fresnel zone inside the Earth that is several hundred kilometers wide. The inversion of integral equations demands care to avoid that noise (errors in the data) is interpreted as small differences in such averages which may require large spatial variations in the Earth's properties. This imposes a fundamental limitation to the resolving power of seismic waves that even finite-frequency theory cannot completely avoid. The theoretically best available horizontal resolution, using the highest observable seismic frequencies, is of the order of hundred kilometers in the lower mantle. In practice, this lower limit is not yet reached because of the limited coverage of the Earth's surface with seismic stations. On land, the coverage can be improved markedly using temporary deployments of seismic arrays, but this is much more difficult and expensive to do in the oceanic domain. Major improvements on a global scale are therefore to be expected only if we solve the problem of oceanic seismometry.

Bibliography

Nolet, G., 2008. *A Breviary of Seismic Tomography*. Cambridge: Cambridge University Press.

Rawlinson, N., Pozgay, S., and Fishwick, S., 2010. Seismic tomography: a window into deep earth. *Physics of the Earth and Planetary Interiors*, **178**, 101–135.

Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years. *Annual Review of Earth and Planetary Sciences*, **31**, 303–328.

Sigloch, K., McQuarrie, N., and Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography. *Nature Geoscience*, **1**, 458–462.

Tromp, J., Tape, C., and Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, **160**, 195–216.

Cross-references

Earth, Density Distribution
Earth's Structure, Core
Earth's Structure, Lower Mantle
Earth's Structure, Upper Mantle
Free Oscillations of the Earth
Gravity Field of the Earth
Inverse Theory, Global Optimization
Inverse Theory, Linear
Inverse Theory, Monte Carlo Method
Mantle Plumes
Plate Driving Forces
Propagation of Elastic Waves: Fundamentals
Seismic Imaging, Overview

SEISMIC VELOCITY-DENSITY RELATIONSHIPS

Kalachand Sain
National Geophysical Research Institute, Council of Scientific & Industrial Research (CSIR), Hyderabad, India

Definition

Seismic velocity is defined as the speed with which a seismic wave travels through a medium and is measured in km/s. The density is the mass per unit volume and is expressed in g/cc.

Types of seismic waves

The *body wave* and the *surface wave* are the main seismic waves. The body waves are the longitudinal (P) and the shear (S) waves in which the particles vibrate parallel and perpendicular to the direction of wave propagation respectively. The velocities of P-wave (V_P) and S-wave (V_S) are related to the density, ρ as

$$V_P = \sqrt{\frac{\kappa + \frac{4}{3}\mu}{\rho}} \quad (1)$$

$$V_S = \sqrt{\frac{\mu}{\rho}} \quad (2)$$

where κ and μ are the bulk and rigidity moduli respectively. It appears that the seismic velocity and density are inversely related, but the denser rocks are characterized by higher velocities compared to lighter rocks because the density also depends on κ and μ. The halite with low density (1.8 g/cc) and high V_P (4.5 km/s) shows an exception. Generally velocities and densities increase with depth. The inversion of this trend also is observed due to pore pressure, fluids, or sediments below volcanics.

General relation between velocity and density

(a) Birch's law – To fit measurements from crustal and mantle rocks, Birch (1961) established a linear relationship between V_P and ρ as

$$V_P = a + b\,\rho \quad (3)$$

where a and b are empirical parameters.

(b) Gardner relation – Gardner et al. (1974) conducted field and laboratory measurements on saturated sedimentary rocks and determined the velocity–density relation as

$$\rho = aV_P^b \quad (4)$$

where $a = 0.31$ and b is 0.25. Major sedimentary rocks fall within a narrow corridor around the Gardner curve. Coals, anhydrite, and salts exhibit large deviations from this trend. The deviation increases with porosity, and becomes higher for lower densities.

Estimation of seismic velocity and density from surface measurements provides important inputs to understand the structure and tectonics of the earth. Since gravity is a potential field, velocity structure is often used to constrain the gravity modeling (Behera et al., 2004) by converting the velocity into density using above formulas. Use of Ludwig et al. (1970) formula to relate V_P with ρ is also very common in crustal studies.

Bibliography

Behera, L., Sain, K., and Reddy, P. R., 2004. Evidence of underplating from seismic and gravity studies in the Mahanadi delta and its tectonic significance. *Journal of Geophysical Research*, **109**, 1–25.

Birch, F., 1961. The velocity of compressional waves in rocks to 10 kilobars. *Journal of Geophysical Research*, **66**, 2199–2224.

Gardner, G. H. F., Gardner, L. W., and Gregory, A. R., 1974. Formation velocity and density – the diagnostic basics for stratigraphic traps. *Geophysics*, **39**, 770–780.

Ludwig, W. J., Nafe, J. E., and Drake, C. L., 1970. Seismic refraction. In Maxwell, A. E. (ed.), *The Sea*. New York: Wiley-Interscience, Vol. 4, Part 1, pp. 53–84.

Cross-references

Body Waves
Surface Waves

SEISMIC VELOCITY-TEMPERATURE RELATIONSHIPS

Kalachand Sain
National Geophysical Research Institute, Council of Scientific & Industrial Research (CSIR), Hyderabad, India

Definition

The seismic velocities of subsurface rocks decrease with temperature but experience opposite effects due to increase in pressure (depth).

Velocity–temperature equation

On an average, the pressure increases at a rate of 30 MPa/km and the temperature raises at a rate of 25°C/km from surface to a few tens of kilometers. Further deep down, this increasing rate of temperature gradient decreases to small values. Since seismic velocity (V) decreases with temperature (T) and increases with pressure (P), we need to know the combined effects of P and T for interpreting seismic velocities at different depths. The generalized relation for the variation of velocity with depth, Z can be expressed as

$$\frac{dV}{dZ} = \left(\frac{\partial V}{\partial P}\right)_T \frac{dP}{dZ} + \left(\frac{\partial V}{\partial T}\right)_P \frac{dT}{dZ} \quad (1)$$

where $\left(\frac{\partial V}{\partial P}\right)_T$ denotes the change in velocity with pressure at constant temperature (isotherm) and $\left(\frac{\partial V}{\partial T}\right)_P$ is the change in velocity with temperature at constant pressure (isobar). $\frac{dP}{dZ}$ and $\frac{dT}{dZ}$ are the vertical pressure and temperature gradients respectively.

To make petrological inferences, the crustal and lithospheric seismic velocities are to be corrected to the experimental reference values (i.e., constant pressure of 100 MPa and room temperature of 20°C). Rybach and Buntebarth (1984) have defined the correction factor (f) to the P-wave velocity (V_P) at a given P and T as

$$V_P(20^\circ\text{C}, 100\text{ MPa}) = V_P(P,T)f = V_P\left[1 + \Delta T\frac{\partial V_P}{V_P \partial T} - \Delta P\frac{\partial V_P}{V_P \partial P}\right] \quad (2)$$

As per Equation 2, the corrections to the field velocities at depths from 1 to 50 km are of the order of ±0.1 km/s.

Velocity variation with temperature for some reservoir rocks

The change in seismic velocities with temperature also depends on the saturation of rock. The results of Wang and Nur (1990) in heavy oil sands and hydrocarbon-saturated

rocks show that both V_P and V_S decrease with increasing temperature. Since V_S is not affected by fluids, the decrease in V_S is due to changes in rock frame and in rock fluid interactions. The V_P in heavy oil sands (Tosaya et al., 1987) shows a dramatic decrease. As temperature increases from 25°C to 125°C, V_P can drop by 35% to almost 90%. Heavy oils are highly viscous and a strong interfacial force exists between oil and rock grains. The viscosity of oil and interfacial force decrease due to rise in temperature, which decreases the rigidity and bulk modulus leading to reduction in seismic velocities. For temperature up to about 150°C, changes in pore fluid properties play dominant role in changing the velocity.

Bibliography

Rybach, L., and Buntebarth, G., 1984. The variation of heat generation, density and seismic velocity with rock type in the continental lithosphere. *Tectonophysics*, **l03**, 335–344.

Tosaya, C., Nur, A., Vo-Thanh, D., and Da Prat, G., 1987. Laboratory seismic method for remote monitoring of thermal EOR. *SPE Reservoir Engineering*, **2**, 238–242.

Wang, Z., and Nur, A., 1990. Wave velocities in hydrocarbon saturated rocks: experimental results. *Geophysics*, **55**, 723–733.

SEISMIC WAVE PROPAGATION IN REAL MEDIA: NUMERICAL MODELING APPROACHES

Ursula Iturrarán-Viveros[1], Francisco J. Sánchez-Sesma[2]
[1]Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, México DF, Mexico
[2]Instituto de Ingeniería, Universidad Nacional Autónoma de México, Ciudad Universitaria, México DF, Mexico

Synonyms

Elastic waves in homogeneous and inhomogeneous media

Definition

The Earth is a complex medium containing heterogeneities and scattering structures over many scales. The influence of heterogeneities on seismic wave propagation is therefore studied intensively by numerical modeling to explain observations. Coming along with the enormous increase of computational power, the development of numerical techniques for the accurate calculation of synthetic seismograms in three-dimensional (3-D) models of the Earth has been the subject of a continuous effort in the last 30 years. This chapter presents a glimpse of several of these numerical methods.

Introduction

In 1969, Keiiti Aki first focused attention on the appearance of continuous wave trains (called coda), in the tail portion of individual seismograms of local earthquakes as a direct evidence of the random heterogeneity of the lithosphere. Models for seismic wave propagation through inhomogeneous elastic media have been developed using deterministic approaches such as model theory for layered media. In 1954, Norman Haskell made a great contribution to theoretical geophysics with his famous paper in which he showed how seismic surface waves could be computed for an Earth modeled by an arbitrary number of plane parallel layers, each one with arbitrarily prescribed physical properties. His work was based on Thomson's work and it is now known as the Thomson–Haskell method. In his pioneering work Claerbout (1968) recognized the solution to the inverse problem of determining the medium from the seismogram. Numerical modeling of seismic wave propagation in an inhomogeneous media is becoming more important in seismic exploration. A wide variety of recent numerical strategies are concerned with treating Earth models of successively increasing complexity. The realistic computational models can include anisotropic media, nonplanar interfaces between layers and blocks, velocity/density/quality-factor gradients inside layers, and often with free-surface topography. In particular, the rheology of the medium should allow for realistic attenuation. We can divide the different approaches to numerically solve the wave equation in complex media into three main classes: *direct methods*, *integral equation methods*, and *ray tracing methods*, see Table 1 and Carcione et al. (2002). In order to solve the wave equation using *direct methods*, the geological model needs to be approximated by a numerical mesh. There are no restrictions about material variability and different rheologies can be implemented. These methods give solution to the wave equation in the time domain. They can be very accurate provided a sufficiently fine grid, but they are computationally more expensive than analytical or ray tracing methods. *Integral equation methods* are based on Huygens' principle. These methods formulate the solution of a problem in terms of values at the domain's boundary. Integral equation methods have an interesting conceptual advantage over direct methods, which is the reduction of one space dimension. Moreover, integral equation methods (in the frequency domain) do not need absorbing boundaries. They match easily the boundary conditions and do not suffer from grid

Seismic Wave Propagation in Real Media: Numerical Modeling Approaches, Table 1 Classification of numerical methods to model seismic wave propagation in real media

Direct methods	Finite differences (FD), pseudospectral (PS), finite element (FE), spectral element method (SEM), finite volume (FV), discontinuous Galerkin (DG), discrete element method (DEM)
Integral equation methods	Boundary element method (BEM), indirect boundary element method (IBEM), fast multipole method (FMM), domain integral methods (DIM)
Asymptotic or ray tracing methods	Gaussian beams

dispersion. However, these methods require the knowledge of Green's function which is limited to a few cases (see Kausel, 2006). *Ray tracing* or *Asymptotic methods* are approximative because they do not take into account the complete wavefield. The wavefield is considered as an ensemble of certain events, each arriving at certain traveltime and having certain amplitude.

Finite-difference method

Among direct methods we have those that solve the strong form of the wave equation (FD, PS) and those that solve the weak form, which are also known as variational methods (such as FE). The strong formulation works directly with the equation of motion and the boundaries are prescribed in their differential form. The weak formulation instead uses an integral form of the equation of motion weighting the error by trial functions. The natural boundary conditions are automatically satisfied in the weak formulation. The basic idea of finite-difference methods is to compute the partial derivatives by an approximation based on Taylor series expansions of functions near the point of interest. Forward and backward time derivatives lead to implicit or explicit schemes. In general, an implicit method is computationally more expensive than an explicit method, due to the requirement of solving large matrix equations. When applying finite differences in the implicit scheme only for 1-D problems, the matrix to be inverted is tridiagonal, which is very convenient in terms of computational cost. However, this is not the case for 2-D problems. Nevertheless, using the method of dimensional splitting, one can replace the multi-dimensional problem by a succession of much simpler one-dimensional problems and as a result it is possible to build tridiagonal matrices and combine their results to compute the desired time derivative. This ingenious approach is described in Bayliss et al. (1986), Mufti (1985) and Emerman et al. (1982). More recently Liu & Sen (2009) have obtained implicit finite-difference formulae derived from fractional expansion of derivatives which form a system of equations that form a tridiagonal matrix. Obviously, the explicit schemes are computationally simpler. Therefore a vast majority of earthquake ground motion modeling and exploration seismology studies use explicit FD schemes. Virieux in a couple of papers (1984) and (1986) for SH and P-SV cases, respectively, was the first to introduce explicit time domain staggered FD in seismology. This method is easy to implement, to introduce different kinds of sources, and to consider viscoelasticity. In addition, the local nature of finite-difference operators makes the method suitable for parallelization. In seismic applications, the following velocity-stress formulation is widely used:

$$\begin{aligned} \frac{\partial \sigma_{ij}}{\partial x_j} + f_i &= \rho \frac{\partial v_i}{\partial t} \\ \frac{\partial \sigma_{ij}}{\partial t} &= c_{ijkl} \frac{\partial e_{kl}}{\partial t} \end{aligned} \tag{1}$$

where the first time derivative of the displacement u is the velocity v, σ_{ij} is the stress, $\partial_t e_{kl} = (v_{k,l} + v_{l,k})/2$, e_{kl} being the strain, c_{ijkl} is the fourth-order stiffness tensor with elastic coefficients for linear elastic solids, and f_i is the ith component of the body force vector $\mathbf{f}$ that can be written in terms of the moment tensor as follows:

$$\mathrm{f} = -\mathbf{M} \cdot \nabla \delta(x - x_s) S(t) \tag{2}$$

where x_s denotes the source position δ is the Dirac delta and $S(t)$ denotes the source time function. Evaluating the second time derivative of Equation 1, i.e., $\ddot{u} = \partial_t v$ (double dot denotes twice time derivative), at times $(n + 1)\Delta t$ and $(n–1)\ \Delta t$ by a Taylor expansion yields

$$\frac{\partial^2 u^n}{\partial t^2} = \frac{1}{(\Delta t)^2} \times \left[u^{n+1} - 2u^n + u^{n-1} - 2 \sum_{l=2}^{L} \frac{(\Delta t)^{2l}}{(2l)!} \frac{\partial^{2l} u^n}{\partial t^{2l}} \right] \tag{3}$$

when applied Equation 3 to numerically approximate spatial and time derivatives in the 1-D scalar wave equation

$$\frac{1}{\beta^2} \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \tag{4}$$

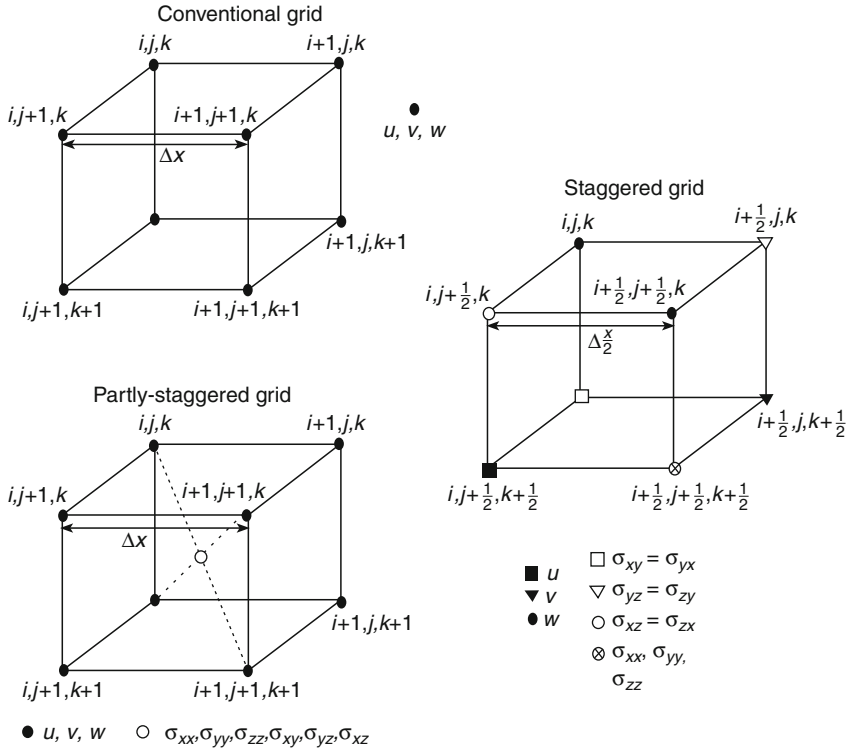
we obtain the Lax-Wendroff scheme

$$\frac{1}{\beta^2} \frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{(\Delta t)^2} = \frac{u_{i+1}^n - 2u_i^n + 2u_{i-1}^n}{(\Delta x)^2} \tag{5}$$

This equation represents the discretization of Equation 4 taking into account that we have a Cartesian system of coordinates (x,y,z) with a fourth-order computational domain taking into account the time, i.e., (x,y,z,t). Considering the spatial-time grid given by a set of discrete points (x_i,y_j,z_k,t_n) given by $x_i = x_0 + i\Delta x$, $y_j = y_0 + j\Delta y$, $z_k = z_0 + k\Delta z$, $t_n = t_0 + n\Delta t$; $i,j,k,n = 1,2,...$ The spatial increments Δx, Δy, and Δz are usually referred to grid spacings, (x_0, y_0, z_0, t_0) are initial values for these variables, while Δt is the time step. The value of a function u at a grid position (x_i, y_j, z_k, t_n) is given by $u_{i,j,k}^n = u(x_i, y_j, z_k, t_n)$. It is possible to have nonuniform grids and the other important consideration is whether all functions are approximated at the same grid positions. The conventional, partly staggered, and staggered grids are illustrated in Figure 1. An important observation is pointed out for finite differences solutions of hyperbolic equations. Accuracy increases with increasing Δt, for central differences, up to a limit defined by the Courant-Friederickson-Lewy stability condition (assuming $\Delta x = \Delta z$, in 2-D):

$$\frac{V_{\max} \Delta t}{\Delta x} \leq \frac{1}{\sqrt{2}} \tag{6}$$

where $V_{\max}$ is the maximum velocity considered in the model.



Seismic Wave Propagation in Real Media: Numerical Modeling Approaches, Figure 1 Spatial grid cells in the conventional, partly staggered, and staggered grids (see Moczo et al., 2007).

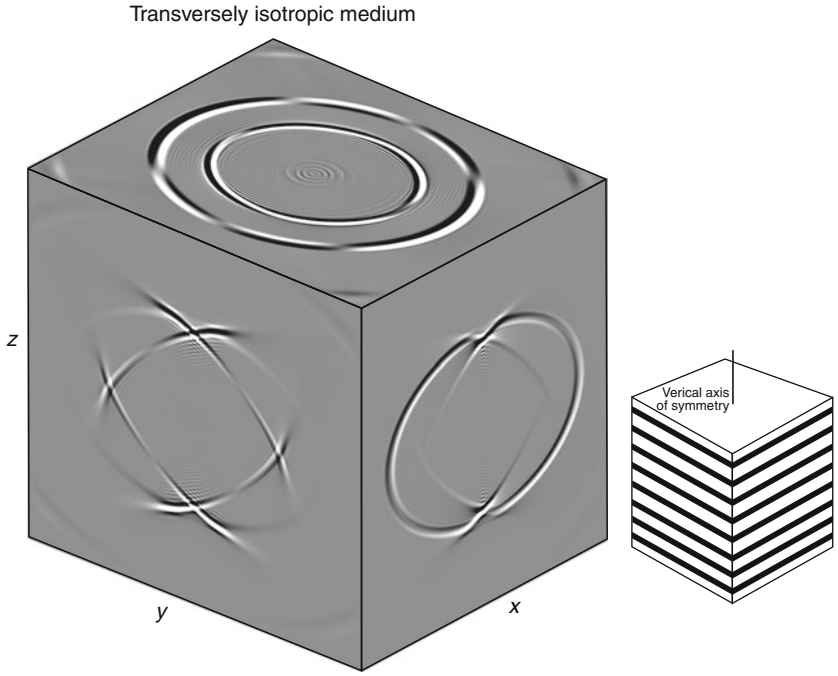
The choice among the different schemes is made to improve accuracy, reduce memory requirements, and to deal with anisotropy media. Recently, the use of the partly staggered grid was upgraded by Saenger et al. (2000) and Saenger & Bohlen (2004). The authors called the grid rotated staggered grid since they obtained the spatial FD operator by the rotation of the standard staggered-grid operator. They used this scheme to successfully simulate media with heterogeneities (cracks, pores, free surface, and anisotropy). Rotated staggered grids have been also used to model planar and nonplanar dynamic rupture models, see Cruz-Atienza et al. (2007). The main limitations of FD are the extensive consumption of computational resources in terms of both core memory and CPU time. However, Furumura and Chen (2005) performed large-scale parallel FD simulations of seismic wave propagation to model strong ground motions during damaging earthquakes. Parallel FD simulation allowed Saito and Furumura (2009) to successfully simulate tsunami generation. A complete review of FD methods can be found in Moczo et al. (2007).

Anisotropy

Anisotropy is caused by structural alignment of elements that are smaller than the wavelength employed for seismic data acquisition. Examples of structures that cause anisotropic effects are fine layering, tectonic grain, and orientation and intensity of cracking. One way to include these effects in the model is to consider that elastic properties at a given point might vary with direction. Therefore, by including anisotropy in the model, alignments of small scale hetegoeneities are captured. The alignment effects might lead to important petrophysical information. To consider anisotropy in a FD simulation one needs to define c_{ijkl} the fourth-order stiffness tensor in Equation 1. For example, to model the effect of layering (see Figure 2) one can consider the following stiffness tensor:

$$C = \begin{bmatrix} c_{11} = 66.6 & c_{12} = 19.7 & c_{13} = 39.4 & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} = 39.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} = 10.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{bmatrix} \tag{7}$$

where $c_{66} = (c_{11} - c_{12})/2$, the density given by $\rho = 2590\ \mathrm{Kg/m^3}$, and the units for the elastic constant are in GPa. The example represents the Mesaverde clay shale (see Thomsen, 1986 and Carcione et al., 1992). This kind of stiffness tensor with these five elastic constants characterizes a medium of class IV according to the classification given by Payton (1983) and it is called transversely



Seismic Wave Propagation in Real Media: Numerical Modeling Approaches, Figure 2 (*Left*). Illustration of a transversely isotropic medium. The example represents the Mesaverde clay shale. The six elastic constants (for the stifness tensor in Equation 1) that generate this medium are given by c_{11} = 66.6 GPa, c_{12} = 19.7 GPa, c_{13} = 39.4 GPa, c_{33} =39.9 GPa, c_{44} = 10.9 GPa., with $c_{66} = (c_{11} - c_{12})/2$ and the density ρ = 2590 Kg/m^3. (*Right*) A layered medium that represents the anisotropic effect.

isotropic medium (since it has the vertical z-axis as a symmetrical axis). Notice that due to the anisotropy, in Figure 2, and the wave fronts do not propagate at the same speed in all directions (see the different wave fronts in plane cuts: xy-plane, xz-plane and yz-plane).

Pseudospectral method

The pseudospectral method (PS) is base mainly on Fourier and Chebyshev differential operators. The spatial derivative is computed in the frequency-wavenumber domain as follows:

$$\frac{\partial u(x,t)}{\partial x} = \frac{1}{2\pi} \int_{-\infty}^{\infty} u(x,t) ike^{ikx} dk \tag{8}$$

see Kreiss and Oliger (1972), Kosloff and Baysal (1982), Furumura et al. (1998). The PS with Fourier has the disadvantage of periodic properties, called wraparound. Both Fourier and Chebyshev are accurate up to the maximum wavenumber of the mesh that corresponds to a spatial wavelength of two grid points (at maximum grid spacing for the Chebyshev operator). Like FD methods, they are also unable to model surface waves with the same accuracy as body waves because of the one-way treatment that needs to be performed in order to implement the free-surface condition (e.g., Carcione, 1994).

Finite element and spectral element methods

The finite element (FE) method is more efficient than FD or PS when dealing with complex geometries (irregular interfaces), heterogeneous media, and handling boundary conditions. This method is also applicable with inelastic constitutive models. Recently, Bielak et al. (2005) have developed an efficient FE-based computational platform for large-scale ground motions. Low approximation orders may lead to a large numerical dispersion, see Marfurt (1984). As a result, mesh refinement is required to reduce numerical dispersion but may lead to a large numerical cost even if parallelization is possible. High-order FE methods became more efficient when the spectral element method (SEM) appeared using Chebyshev polynomials as interpolating functions, see Seriani et al. (1992), Priolo et al. (1994). Komatitsch and Villotte (1998), and Komatitsch and Tromp (1999). This method introduced the use of the Lagrange polynomials as interpolating functions which produces a diagonal mass matrix when using the Gauss-Lobatto-Legendre quadrature.

Discontinuous Galerkin method

The Discontinuous Galerkin method (DG) is a technique that uses discontinuous basis functions to formulate a Galerkin approximation. Given a mesh of the analysis domain, the DG method approximates the solution within each element by a function from a low-dimensional vector

space of functions, e.g., as a linear combination of basis functions like polynomials. For a pair of adjacent mesh elements, the approximate solution computed in the interior of the elements does not have to agree on their common boundary. Some of the advantages of this method are:

- It can sharply capture solution discontinuities relative to a computational mesh.
- It simplifies adaptation since inter-element continuity is neither required for mesh refinement and coarsening.
- It conserves the appropriate physical quantities (e.g., mass, momentum, and energy) on an element-by-element basis.
- It can handle problems in complex geometries to high order.
- Regardless of order, it has a simple communication pattern to elements sharing a common face that simplifies parallel computation. With a discontinuous basis, the DG method produces more unknowns for a given order of accuracy than traditional finite element or finite volume methods, which may lead to some inefficiency. The DG method is harder when applied to unstructured meshes; in particular, it is harder to formulate limiting strategies to reduce spurious oscillations when high-order methods are used. Nevertheless, Käser & Dumbster (2006) have successfully implemented the method for elastic wave propagation on unstructured meshes.

For an introduction to the key ideas, basic analysis, and efficient implementation of discontinuous Galerkin methods and a review of the state of the art, we refer the reader to the book by Hesthaven & Warburton (2008).

Integral equation methods

Boundary element and indirect boundary element methods (BEM & IBEM)

Among Integral equation methods it is the boundary element method. The formulation of direct boundary integral equation method (BIE) in elastodynamics is back to the pioneering work of Somigliana (1886). Considering a volume of elastic material V bounded by a surface S, the displacement field u_i at a point $\mathbf{x}$ and time t can be expressed as a function of the values of the displacement and traction t_j along the boundary through Somigliana's representation theorem:

$$cu_i(\xi,t) = \int_0^t \mathrm{d}\tau \int_s \left[t_j(\xi,\tau)G_{ij}(\xi,t,x,\tau) - u_j(\xi,\tau)\Sigma_{jik}(\xi,t,x,\tau)n_k(\xi)\right]\mathrm{d}S(\xi) \tag{9}$$

where $G_{ji}(\xi,t,\mathbf{x},\tau)$ and $\Sigma_{jik}(\xi,t,\mathbf{x},\tau)$ are the responses in terms of displacement and stress, time t and point ξ of an infinite homogeneous medium to a unit force impulse, applied at time τ at point $\mathbf{x}$ in the direction i, $n_k(\xi)$ is the normal boundary pointing outside V. The volumetric body sources are assumed to be null, but if present, their contribution can be easily added as follows:

$$cu_i(\xi,t) = \int_0^t \mathrm{d}\tau \int_s \left[t_j(\xi,\tau)G_{ij}(\xi,t,x,\tau) - u_j(\xi,\tau)\Sigma_{jik}(\xi,t,x,\tau)n_k(\xi)\right]\mathrm{d}S(\xi) + \int_v^t \mathrm{d}\tau \int_V G_{ij}(\xi,t,x,\tau)f_i(x,\tau)\mathrm{d}V \tag{10}$$

where

$$c = \begin{cases} 1, & \xi \in \mathrm{V} \\ 1/2, & \xi \in \mathrm{S} \\ 0, & \xi \notin \mathrm{V} \end{cases} \tag{11}$$

and assuming that S has a smooth boundaries. The values of c come from the volume integration of the product $\delta(\mathbf{x}-\xi, t\text{-}\tau)\,\mathbf{u}_j(\xi,\tau)$ and the value accounts for the integration of the Dirac's delta and its position. The Green's function for the homogeneous, isotropic, elastic media can be found in Aki & Richards (1980).

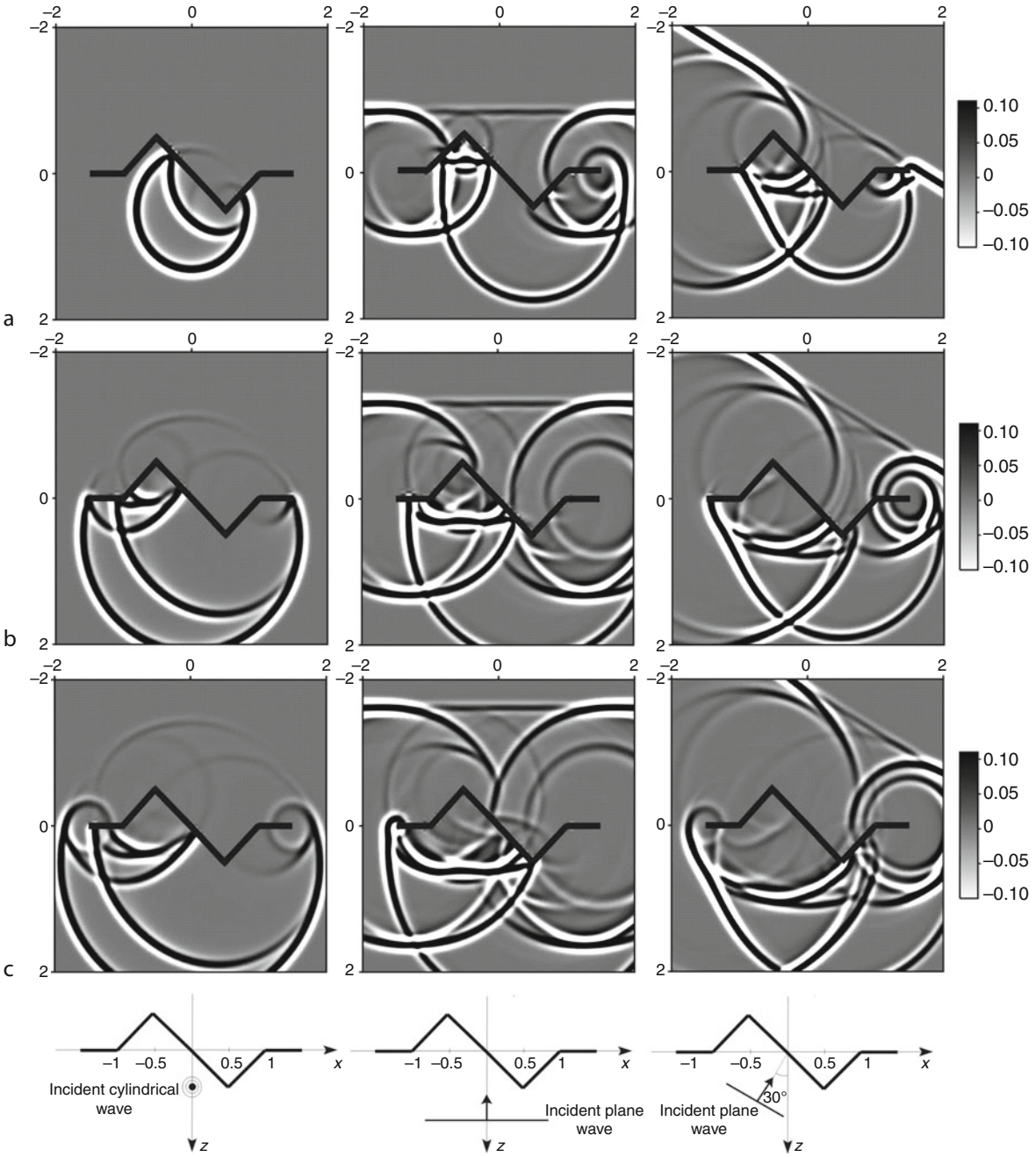
The indirect formulation of the elastodynamic problem expresses the wavefield as an integral over the boundary of elementary source radiations as follows:

$$u_i(x,t) = \int_0^t \mathrm{d}\tau \int_s \phi_j(\xi,\tau)G_{ij}(x,t,\xi,\tau)\mathrm{d}S(\xi) \tag{12}$$

where $\phi_j(\xi,\tau)$ denotes τ the force density distribution applied at time t at point ξ of the surface S. In this equation we have assumed that the volumic forces are null. If present, their contribution needs to be added. Equation 10 can be seen as the mathematical transcription of Huygens' principle, which states that each point of diffracting boundary acts as a secondary source of wave radiation. This principle applies to any wavefront as well. Equation 12 is also the mathematical expression of the exploding reflector concept widely used in seismic exploration (see Claerbout, 1985). The fictitious distribution ϕ_j is an intermediate variable which needs to be solved in the boundary integral equations before the wavefield can be computed. This is the reason why this formulation is called indirect.

The discretization of the boundaries leads to the discretization of the BIE, and the application of boundary conditions transforms BIE into a system of linear equations which in general is not symmetric. The resolution of this system is usually done implicitly and in the frequency domain. Comprehensive intorduction of integral equation methods can be found in the books by Bonnet (1999), Dominguez (1993), and Manolis & Beskos (1988). In seismology, the BEM is used to study the effect of irregular topography on earthquake ground motion, in

2-D some other works deal with 3-D problems. This method has been also used to simulate elastic wave propagation in media with cracks, for example, on Figure 3 we have an irregular crack strike by a cylindrical SH incident wave (first column) and the same crack strike by a plane SH incident wave with 0^o and 30^o, measured with respect to the z-axis (second and third columns, respectively), see Iturrarán-Viveros et al., 2005; Pointer et al., 1998. A complete review on the use of BEM for seismic problems can be found in Bouchon and Sánchez-Sesma (2007).



Seismic Wave Propagation in Real Media: Numerical Modeling Approaches, Figure 3 Snapshots of displacements for: (*left*) A cylindrical SH wave striking an irregular crack. (*middle*) A plane SH wave with incident angle 0^o striking an irregular crack. (*right*) A plane SH wave with incident angle 30^o striking an irregular crack. The set of snapshots are taken at (**a**) $t = 1.25$ s (**b**) $t = 1.71$ s and (**c**) $t = 2.03$ s.

Fast multipole method in elastodynamics

Two types of fast multipole method (FMM) are available for elastodynamics in the frequency domain. The first one is the low frequency FMM. As for static cases, the complexity of this algorithm is $O(N)$. This complexity is due to the fact that the wavelength is much larger than the domain size. On the other hand, if the wavelength is shorter than the geometrical feature, the complexity of low frequency FMM increases to $O(N^2)$ and the method is not efficient any longer. For this reason, computational efficiency of fast BEMs in the mid-frequency regime is enhanced by using the so-called diagonal form for the Helmholtz Green's function, proposed by Rokhlin (1990), with a complexity of O ($N \log N$). The upper limit stems from the fact that the size N becomes intractable at high frequencies, but the diagonal form also breaks down at very low frequencies and must be replaced with other types of expansions. The method for low frequencies was developed by Fujiwara (1998). The first 3-D implementation was proposed by Fujiwara (2000) using a multi-level and diagonal form. The author presents some low frequency seismic oriented examples. More recently Chaillat et al. (2009) also addressed problems in 3-D. Yoshida (2001) proposed a low frequency FMM for crack problems in 3-D.

Analytic solutions for nonhomogeneous media

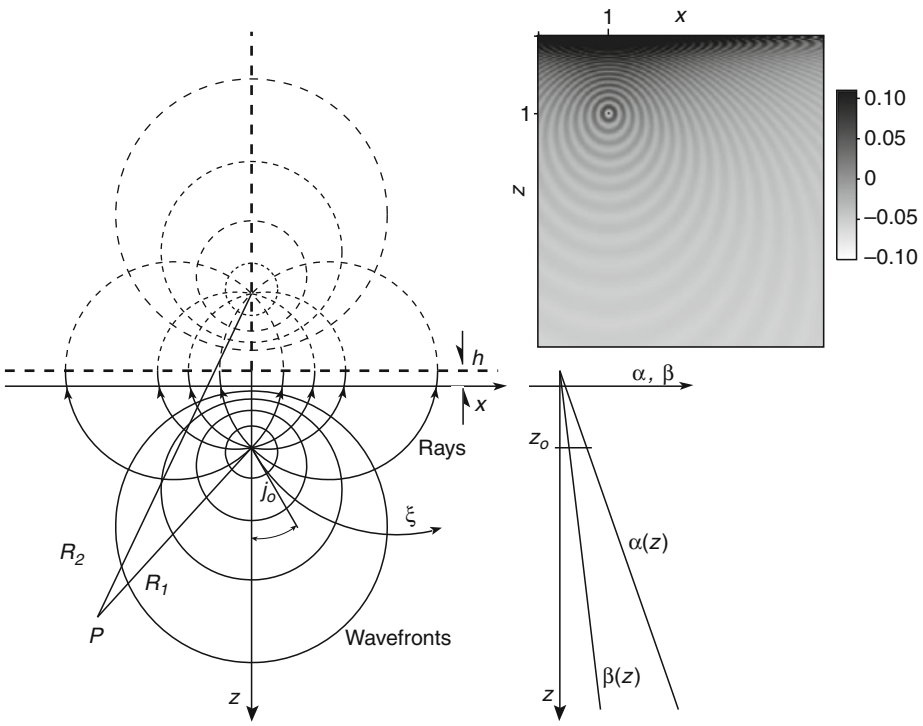
Since analytic solutions help to assess numerical methods, some attempts to develop analytic solutions for particular cases of mildly anisotropy media are briefly described. In Sánchez-Sesma et al. (2001) authors developed an approximate analytical formula for elastic 2-D Green's functions for a constant-gradient propagation velocity medium. These solutions correspond to unit line forces per unit length: the antiplane SH line source and the in-plane P-SV line sources, respectively. They are based on the asymptotic ray theory and account for both near-source effects and low frequencies. The orthogonal bipolar coordinates are depicted in Figure 4.

The approximate solution for the 2-D scalar case is given by

$$G_{22}(\omega,\tau) \approx \Lambda \frac{1}{4\mu_0} H_0^{(1)}(\omega\tau) \tag{13}$$

where Λ is defined as follows:

$$\Lambda = \left(\frac{1+\gamma z_0}{1+\gamma z}\right)^{\frac{1+n}{2}} \left[2\ln\left(\frac{R_2+R_1}{R_2-R_1}\right)\frac{(z_0+h)(z+h)}{R_1 R_2}\right]^{1/2} \tag{14}$$



Seismic Wave Propagation in Real Media: Numerical Modeling Approaches, Figure 4 (*Left*). Schematic illustration of the bipolar coordinate system. Both rays and wave fronts are circular. Rays depend upon take-off angle j_0. $t = \xi t_0$, where $t_0 = h/\beta(0)$ and ξ is one of the bipolar coordinates (the other is j_0). The distances R_1 and R_2 from the poles to point P are displayed. The plane for which $z = -h$ corresponds to null propagation velocities. (*Top right*) This solution is used to compute the Green's function for the scalar case in the time domain with a source located at (1,1), $\beta(0)$ = 500 m/s, and $\beta(z_{max})$ = 3,000 m/s. Wave fronts depend on traveltime.

where z_0 = depth of the source, β_0 = shear wave velocity at the source level, $\beta(0)$ = shear wave velocity at $z = 0$, ρ_0 = mass density at source level, $\gamma = 1/h$, h = distance from depth $z = 0$ to the level where the propagation velocity is null ($z = -h$), $j_0 = \sin^{-1}[2x(z_0 + h)/R_1R_2]$ = take-off angle, $R_1 = (x^2 + (z-z_0))^{1/2}$, and $R_2 = (x^2 + (z + z_0 + 2h)^2)^{1/2}$. The traveltime t is given by

$$\tau = t_0 \ln\left(\frac{R_2 + R_1}{R_2 - R_1}\right) = 2t_0 \tanh^{-1}\left(\frac{R_1}{R_2}\right) = t_0\xi \qquad (15)$$

In Watanabe & Payton (2002, 2006) Green's functions are obtained exactly for radially symmetric waves and for transient and time-harmonic SH-waves in inhomogeneous anisotropic elastic solids, respectively. Since time-harmonic conditions have been assumed, viscoelastic material behavior can be captured through the introduction of complex-valued wave speeds.

Equivalent media theories

When one wishes to model highly heterogeneous media, one possibility is to make the continuum hypothesis (see Christensen, 1979; Hudson, 1991). A real material with its atomic and molecular structure is replaced by an equivalent continuum model, which remains a continuum no matter how much it is subdivided. The macroscopic mechanical properties of the material remain unchanged. This hypothesis works as long as the scale of observation is much larger than the scale-length of the molecular structure (Rayleigh scattering regime). Micro fractures, pores, and other heterogeneities with uniform statistical distributions can be replaced by an equivalent or effective medium provided that the scale of the observation is much larger than the scale of the heterogeneities. This is an attractive approach in exploration seismology since the scale of the observation, or seismic wavelengths, is much larger than the scale length of the heterogeneities present in the subsurface. The concept of equivalent media clearly implies that, if the stress, strain or displacements are measured on a large enough scale, the values obtained and the relationship between them will be that of homogeneous continuum. This implies some kind of spatial averaging process. Nevertheless, scales much smaller than the minimum wavelength (used in the numerical model) are present in the earth model. Therefore effective media theories allow to average small scales of the original medium without losing the accuracy of the wavefield computation that are needed. In Capdeville and Marigo (2007), authors applied two-scale homogenization technique to the wave equation in layered media. The order 0 homogenization provides the same result as the one given by Backus (1962). The order 0 is not enough to obtain an accurate solution especially for surface waves. Higher order homogenization terms (up to 2) allowed them to obtain accurate surface waves. Lately Capdeville et al. (2010) proposed original nonperiodic homogeneization technique in 1-D. The extensions to 2-D and 3-D are foreseen.

Asymptotic ray tracing methods

An accurate estimation of traveltime is needed to map heterogeneities in the Earth's subsurface. Traveltime computation schemes fall within one of these two categories: First-arrival traveltime, and ray-theory based traveltime. The first-arrival traveltime corresponds to the first arrival of the complete wavefield at a specified receiver position and it is not a function of the type of wave (e.g., head wave, direct wave). Most of the schemes, which estimate first-arrival traveltimes, are based on the solution of eikonal equation. Vidale (1988, 1990) proposed different versions of finite-difference solution of the eikonal equation along expanding square (in 2-D) and along an expanding cube (in 3-D) to compute first-arrival traveltimes in isotropic media. Schneider et al. (1992) proposed a method based on Fermat's principle for traveltime computation in isotropic media that uses a local ray-trace solution of the eikonal equation. Ray tracing can be classified into two categories: (1) Initial-value ray tracing, and (2) Boundary value ray tracing. In initial-value ray tracing, a fan of rays is shot from the source in the model and some kind of extrapolation scheme is used to estimate the traveltime at a particular point in the model. Paraxial extrapolation (Cêrvený, 1985) is one of the most popular methods of traveltime extrapolation. An excellent review on ray theory can be found in (Cêrvený, 2001).

Poroelastic media

In addition to complex geometries, different rheologies and anisotropy, current numerical techniques are concentrating efforts to include porosity and fluid saturation in their models. Reservoir-rocks can be modeled as a porous solid matrix where the pore space is fully saturated with a pore fluid. The analysis of wave propagation in fluid saturated porous media is different in nature than that of elastic wave propagation. Biot's model is generally used to describe this effect (Biot 1956; Biot 1962). Two approaches exist to derive the poroelastic equations of motion and the constitutive laws. One method is to use homogenization methods. The laws to describe the medium are first derived on the microscopic scale of porosity. Then they are passed to macroscopic scale by homogenization. Pride et al. (1992) used an averaging procedure to derive the poroelastic equations similar to Biot (1956) and Biot (1962). We refer the reader to the book by Carcione (2007).

Boundaries

When modeling the propagation of elastic waves in any medium, the waves that reach the neighborhood of the boundaries will create reflected waves. This a nonphysical phenomena that we want to avoid. On top of the model we would like to impose a free surface. Therefore, for the top boundary, we need a reflected boundary and for the other three boundaries (in 2-D) we need a nonreflecting absorbing boundaries. In most of the numerical schemes described here, there are boundary

conditions to be imposed or it is necessary to implement absorbing boundaries conditions to eliminate spurious reflections due to the finite computational domain. The free-surface boundary condition is implemented by adding fictitious line of grid nodes next to the top boundary and use the one-side differences to approximate normal derivatives and centered difference to approximate tangencial derivatives, see Kelly et al. (1976) and Graves (1996). The simplest and more effective absorbing boundary conditions consist of a sponge along the artificially truncated edges of the domain. Clayton and Enquist (1977) proposed a kind of absorbing boundary conditions based on replacing the wave equation in the boundary region by one-way wave equation that do not allow energy to propagate from the boundaries to the numerical domain. An alternative scheme suggested by Cerjan et al. (1985) is based on gradual reduction of the amplitudes in a strip of nodes along the boundaries for the mesh. Bérenger (1994, 1996) first introduced Perfectly Matched Layers (PMLs) for electromagnetism. The main idea is to define a selective attenuation of the fields propagating in one prescribed direction. PMLs for elastic wave propagation were developed by Basu and Chopra (2003, 2004), Komatitsch and Martin (2007). The velocity-stress formulation of PMLs for elastic wave equations has been introduced by Collino and Tsogka (2001). It was applied to viscoelastic (Martin & Komatitsch, 2009) or poroelastic (Zeng et al., 2001; Martin et al., 2008) media. Recently Meza-Fajardo and Papageorgiou (2008) demonstrated that for any anisotropic media, the classic PML exhibits instabilities. In addition, they presented a generalization of the classical Perfectly Matched Layer (PML) (called multiaxial Perfectly Matched Layer M-PML) to a medium in which damping profiles are specified in more than one direction.

Summary

Modeling the Earth is a challenging task. With the increase of computational power, direct modeling techniques start to open the door to inversion problems. All numerical methods share at the same time their own disadvantages and weaknesses (high computational costs and storage requirements, poor accuracy, nonstructured grids, hard implementations, system matrix full) with the agreeable flavor of equivalent advantages (reduction of one dimension of the problem, system matrix diagonal or sparse, no need for absorbing boundaries, easy implementation, easy to model topographies or complex geometries). The nature of the problem gives us the appropriate guide to choose one method over others. In this article we briefly cover some numerical techniques to model elastic waves in heterogeneous media. However, the literature on all available methods is very extensive and we include only some references.

Acknowledgments

This work was partially supported by DGAPA-UNAM Project IN121709, Mexico.

Bibliography

Aki, K., and Richards, P. G., 1980. *Quantitative Seismology*. New York: WH Freeman.

Backus, G., 1962. Long-wave elastic anisotropy produced by horizontal layering. *Journal of Geophysical Research*, **67**(11), 4427–4440.

Basu, U., and Chopra, A. K., 2003. Perfectly matched layers for time-harmonic elastodynamics of unbounded domains: theory and finite-element implementation. *Computer Methods in Applied Mechanics and Engineering*, **192**, 1337–1375.

Basu, U., and Chopra, A. K., 2004. Perfectly matched layers for transient elastodynamics of unbounded domains. *International Journal for Numerical Methods in Engineering*, **59**, 1039–1074.

Bayliss, A., Jordan, K. E., LeMesurier, B. J., and Turkel, E., 1986. A fourth-order accurate finite-difference scheme for the computation of elastic waves. *Bulletin of the Seismological Society Of America*, **76**, 1115–1132.

Bérenger, J. P., 1994. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, **114**, 185–200.

Bérenger, J. P., 1996. Three-dimensional perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, **127**, 363–379.

Bielak, J., Ghattas, O., and Kim, E. J., 2005. Parallel octree-based finite element method for large-scale earthquake ground motion simulation. *Computer modeling in engineering and sciences*, **10**, 99–112.

Biot, M., and Biot, M., 1956. Theory of propagation of elastic waves in a fluid saturated porous solid. i-low frequency range. *The Journal of the Acoustical Society of America*, **28**, 168–178.

Biot, M., 1962. Mechanics of deformation and acoustic propagation in porous solids. *Journal of Applied Physics*, **33**, 1482–1498.

Bonnet, M., 1999. *Boundary Integral Equation Method for Solids and Fluids*. Chichester: Wiley.

Bouchon, M., and Sánchez-Sesma, F. J., 2007. Boundary integral equations and boundary element methods in elastodynamics. Advances in wave propagation in heterogeneous earth. In Wu, R.-S., Maupin, V., and Dmowska, R. (eds.), *Advances in Geophysics*. New York: Elsevier-Academic, Vol. 48, pp. 157–190.

Capdeville, Y., Guillot, L., and Marigo, J.-J., 2010. 1-D non periodic homogenization for the wave equation. *Geophysical Journal International*, **181**, 897–910.

Capdeville, Y., and Marigo, J.-J., 2007. Second order homogenization of the elastic wave equation for non-periodic layered media. *Geophysical Journal International*, **170**, 823–838.

Carcione, J. M. D., Kosloff, A. B., and Seriani, G., 1992. A spectral scheme for wave propagation simulation in 3-D elastic-anisotropic media. *Geophysics*, **57**(12), 1593–1607.

Carcione, J. M., 1994. The wave equation in generalized coordinates. *Geophysics*, **59**, 1911–1919.

Carcione, J. M., Herman, G. C., and ten Kroode, A. P. E., 2002. Seismic modeling. *Geophysics*, **67**(4), 1304–1325.

Carcione, J. M., 2007. *Wave Field in Real Media. Wave Propagation in Anisotropic, Anelastic, Propus and Electromagnetic Media*. Netherlands: Elsevier. Handbook of Geophysical Exploration.

Cerjan, C., Kosloff, R., and Reshef, M., 1985. A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, **50**, 705–708.

Cêrvený, V., 2001. *Seismic Ray Theory.* Cambridge: Cambridge University Press.

Cêrvený, V.,1985. The application of ray tracing to the numerical modeling of seismic wavefields in complex structures. In Helbig, K., and Treitel, S., *Handbook of Geophysical Exploration*. London: Geophysical Press, Vol. 15A, pp 1–124.

Chaillat, S., Bonnet, M., and Semblat, J. F., 2009. A new fast multi-domain BEM to model seismic wave propagation and

amplification in 3D geological structures. *Geophysical Journal International*, **177**, 509–531.
Claerbout, J. F., 1968. Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, **33**(2), 264–269.
Claerbout, 1985. *Imagining the Earth's Interior*. Blackwell Scientific Publication.
Clayton, R., and Enquist, B., 1977. Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the Seismological Society of America*, **67**(6), 1529–1540.
Christensen, R. M., 1979. *Mechanics of Composite Materials*. New York: Wiley.
Cockburn B., Karniadakis, G., and Shu C., 2000. Discontinuous Galerkin methods: theory, computation and applications, Vol. 11 of Lecture Notes in Computational Science and Engineering. Springer.
Collino, F., and Tsogka, C., 2001. Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. *Geophysics*, **66**, 294–307.
Cruz-Atienza, V. M., Virieux, J., and Aochi, H., 2007. 3D Finite-difference dynamic-rupture modelling along non-planar faults. *Geophysics*, **72**, SM123–SM137.
Emerman, H. S., Schmidt, W., and Stephen, F. L. A., 1982. An implicit finite-difference formulation of the elastic wave equation. *Geophysics*, **47**(11), 1521–1526.
Frankel, A., and Leith, W., 1992. Evaluation of topographic effects on it P and it S waves of explosions at the northern Novaya Zemlya test site using 3-D numerical simulations. *Geophysical Research Letters*, **19**, 1887–1890.
Dominguez, J., 1993. *Boundary Elements in Dynamics*. London/ New York: Elsevier Applied Science.
Frankel, A., and Vidale, J., 1992. A three-dimensional simulation of seismic waves in the Santa Clara Valley, California, from a Loma Prieta Aftershock. *Bulletin of the Seismological Society Of America*, **82**, 2045–2074.
Fujiwara, H., 1998. The fast multipole method for the integral equations of seismic scattering problems. *Geophysical Journal International*, **133**, 773–782.
Fujiwara, H., 2000. The fastmultipolemethod for solving integral equations of three-dimensional topography and basin problems. *Geophysical Journal International*, **140**, 198–210.
Furumura, T., and Kennett, B. L. N., 1997. On the nature of regional phases – II On the effect of crustal barriers. *Geophysical Journal International*, **129**, 221–234.
Furumura, T., and Chen, L., 2005. Parallel simulation of strong ground motions during recent and historical damaging earthquakes in Tokyo, Japan. *Parallel Computing*, **31**(2), 149–165.
Furumura, T., Kennett, B. L. N., and Takenaka, H., 1998. Parallel 3-D pseudospectral simulation of seismic wave propagation. *Geophysics*, **63**, 279–288.
Graves, R. W., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society Of America*, **86**, 1091–1106.
Hesthaven, J. S., and Warburton, T., 2008. *Nodal Discontinous Galerkin Methods Texts in Applied Mathematics*. New York: Springer Verlag.
Higdon, R. L., 1991. Absorbing boundary-conditions for elastic-waves. *Geophysics*, **56**, 231–241.
Hudson, J. A., 1991. Overall properties of heterogeneous material. *Geophysical Journal International*, **107**, 505–511.
Iturrarán-Viveros, U., Vai, R., and Sánchez-Sesma, F. J., 2005. Scattering of elastic waves by a 2-D crack using the Indirect Boundary Element Method (IBEM). *Geophysical Journal International*, **162**, 927–934.
Kausel, E., 2006. *Fundamental Solutions in Elastodynamics. A Compendium*. New York: Cambridge University Press.
Käser, M., and Dumbster, M., 2006. An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes – I. The two-dimensional isotropic case with external source terms. *Geophysical Journal International*, **166**, 855–877.
Kelly, K. R., Ward, R., Treitel, W., and Alford, R. M., 1976. Synthetic seismograms: A finite difference approach. *Geophysics*, **41**, 2–27.
Komatitsch, D., and Vilotte, J.-P., 1998. The Spectral Element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society Of America*, **88**, 368–392.
Komatitsch, D., and Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation. *Geophysical Journal International*, **139**, 806–822.
Komatitsch, D., and Martin, R., 2007. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, **72**, SM155–SM167.
Kosloff, D., and Baysal, E., 1982. Forward modeling by the Fourier method. *Geophysics*, **47**, 1402–1412.
Kreiss, H.-O., and Oliger, J., 1972. Comparison of accurate methods for the integration of hyperbolic equations. *Tellus*, **24**, 199–215.
Liu, Y., and Sen, M. K., 2009. A practical implicit finite-difference method: examples from seismic modelling. *Journal of Geophysics and Engineering*, **6**, 31–249.
Marfurt, K. J., 1984. Accuracy of finite-difference and finite element modeling of the scalar and elastic wave equations. *Geophysics*, **49**, 533–549.
Martin, R., Komatitsch, D., and Ezziani, A., 2008. An unsplit convolutional perfectly matched layer improved at grazing incidence for seismic wave equation in poroelastic media. *Geophysics*, **73**(4), T51–T61.
Martin, R., and Komatitsch, D., 2009. An unsplit convolutional perfectly matched layer technique improved at grazing incidence for the viscoelastic wave equation. *Geophysical Journal International*, **179**(1), 333–344.
Manolis, G. D., and Beskos, D. E., 1988. *Boundary Element Methods in Elastodynamics*. London: Unwin Hyman.
Manolis, G. D., and Shaw, R. P., 1996. Green's function for the vector wave equation in midly heterogeneous continum. *Wave Motion*, **24**, 59–83.
Meza-Fajardo, K., and Papgeorgiou, A. S., 2008. A nonconvolutional, split-field, perfectly matched layer for wave propagation in isotropic and anisotropic elastic media: Stability analysis. *Bulletin of the Seismological Society Of America*, **98**(4), 1811–1836.
Moczo, P., Robertsson, J. O. A., and Eisner, L., 2007. The finite-difference time-domain method for modeling of seismic wave propagation. *Advances in Geophysics*, **48**, 421–516.
Mufti, I. R., 1985. Seismic modeling in the implicit mode. *Geophysical Prospecting*, **33**, 619–656.
Olsen, K. B., Archuleta, R. J., and Matarese, J. R., 1995. Magnitude7.75 earthquake on the San Andreas fault: three-dimensional ground motion in Los Angeles. *Science*, **270**, 1628–1632.
Payton, R. G., 1983. *Elastic Wave Propagation in Transversely Isotropic Media*. Hague: Mertinus Nijhoff.
Pointer, D., Liu, E., and Hudson, J., 1998. Numerical modeling of seismic waves generated by hydrofractures: application of the indirect boundary element method. *Geophysical Journal International*, **135**, 289–303.
Pride, S., Gangi, A., and Morgan, F., 1992. Deriving the equations of motion for porous isotropic media. *The Journal of the Acoustical Society of America*, **92**(6), 3278–3290.
Priolo, E., Carcione, J. M., and Seriani, G., 1994. Numerical simulation of interface waves by high-order spectral modeling techniques. *The Journal of the Acoustical Society of America*, **95**, 681–693.

Rokhlin, V., 1990. Rapid solution of integral equations of scattering theory in two dimensions. *J. Comp. Phys.*, **86**, 414–439.
Richtmyer, R. D., and Morton, K. W., 1967. *Difference Methods for Initial-Value Problems*, 2nd edn. Nuw York: Wiley.
Thomsen, L., 1986. Weak elastic anisotropy. *Geophysics*, **51**, 1954–1966.
Saenger, E. H., Gold, N., and Shapiro, S. A., 2000. Modeling the propagation of elastic waves using a modified finite-difference grid. *Wave Motion*, **31**, 77–92.
Saenger, E. H., and Bohlen, T., 2004. Finite-difference modeling of viscoelastic and anisotropic wave propagation using the rotated staggered grid. *Geophysics*, **69**, 583–591.
Saito, T., and Furumura, T., 2009. Three-dimensional tsunami generation simulation due to sea-bottom deformation and its interpretation based on the linear theory. *Geophysical Journal International*, **178**, 877–888.
Sánchez-Sesma, F. J., Madariaga, R., and Irikura, K., 2001. An approximate elastic two-dimensional Green's function for a constant-gradient medium. *Geophysical Journal International*, **146**, 237–248.
Schneider, W. A., Ranzinger, K. A., Balch, A. H., and Kruse, C., 1992. A dynamic programming approach to first arrival traveltime computation in media with arbitrary distributed velocities. *Geophysics*, **57**, 39–50.
Somigliana, C., 1886. Sopra l'equilibrio di un corpo elastico isotrope. *Nuovo Cimento*, **20**, 181–185.
Wu, R-S., Maupin V., and Dmowska, R., Seriani G., Priolo, E., Carcione, J. M., and Padovani, E., 1992. High-order spectral element method for elastic wave modelling. Advances in Wave Propagation in Heterogeneous Earth. SEG 62 Annual meeting, Expanded Abstracts, pp. 1285–1288.
Vidale, J. E., 1988. Finite-difference calculation of travel times. *Bulletin of the Seismological Society Of America*, **78**, 2062–2076.
Vidale, J. E., 1990. Finite-difference calculation of traveltimes in three dimensions. *Geophysics*, **55**, 521–526.
Virieux, J., 1986. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, **51**, 889–901.
Virieux, J., 1984. SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, **49**, 1933–1957.
Watanabe, K., and Payton, R., 2002. Green's function for SH-waves in a cylindrical monoclinic material. *Journal of the Mechanics and Physics of Solids*, **50**, 2425–2439.
Watanabe, K., and Payton, R., 2006. Green's function for radially symmetric waves in inhomogeneous anisotropic solids. *Zeitschrift fur Angewandte Mathematik und Mechanik*, **86**, 642–648.
Yoshida, K. I., 2001. *Application of Fast Multipole Method to Boundary Integral Equation Method*. Ph. D thesis, Kyoto University.
Zeng, Y. Q., He, J. Q., and Liu, Q. H., 2001. The application of the perfectly matched layer in numerical modeling of wave propagation in poroelastic media. *Geophysics*, **66**(4), 1258–1266.

Cross-references

Body Waves
Earthquake Rupture: Inverse Problem
Earthquakes, Source Theory
Poroelasticity
Propagation of Elastic Waves: Fundamentals
Seismogram Interpretation
Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting
Statistical Seismology
Surface Waves
Tsunami

SEISMIC WAVES, SCATTERING

Ludovic Margerin
Institut de Recherche en Astrophysique et Planétologie, Observatoire Midi-Pyrénées/C.N.R.S., Toulouse, France

Definition

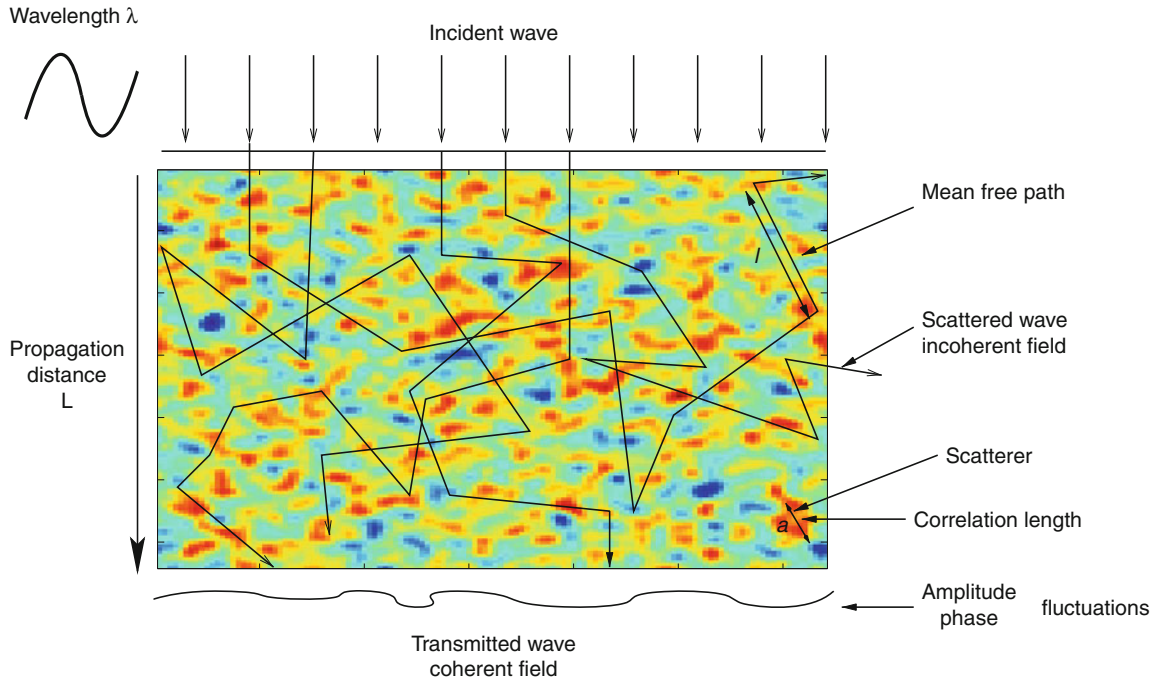
Scattering of seismic waves is best defined by reference to a laterally homogeneous or slowly varying medium where the wave fronts can be perfectly tracked and the propagation of waves can be successfully described by geometrical methods such as ray theory. In the presence of obstacles, or lateral variations of elastic parameters, wave fronts are distorted and seismic energy can be deflected in all possible directions: a phenomenon known as wave scattering. Specific to seismic waves are the possible mode conversions between various polarizations: compressional, shear, and Rayleigh waves.

Introduction

The scattering of seismic waves is intimately related to the heterogeneous nature of Earth materials on a variety of spatial scales. Well-logs data or geological maps offer direct but limited access to the nature of heterogeneities in the Earth. As elastic waves can propagate all the way through the Earth, they constitute the primary tool to sound deep heterogeneities. While the pioneering studies of seismic scattering were largely focused on the lithosphere, the presence of scatterers in the mantle and solid core has been confirmed in a number of recent studies (see Shearer and Earle, 2008, for a review). Hence, the field of seismic scattering has become of interest in global seismology too. For a thorough review of pre-1990s works in seismic scattering, we refer the reader to Wu (1989). A comprehensive reference on the topic is the book by Sato and Fehler (1998). Reviews of recent methodological and observational developments are given in Sato et al. (2008).

Statistical description of the Earth

To introduce the various scale lengths that enter in the modeling of seismic scattering, it is convenient to consider a thought-experiment such as the one shown in Figure 1. We imagine an incident plane wave propagating through a heterogeneous medium. As evidenced by the analysis of well-logs (see for instance Wu et al., 1994), the density and the elastic constants fluctuate in the Earth in a very complicated way. In such a situation, it is convenient to think of the Earth as a particular realization of an ensemble of random media sharing similar statistical properties. The goal of the stochastic imaging methods is to infer some information on the nature of the randomness from the statistical fluctuations of the wavefield. This is a difficult task because the statistical properties of the Earth are nonhomogeneous: they are expected to depend on the geological setting and on the depth. In addition, in the case



Seismic Waves, Scattering, Figure 1 Basic concepts of seismic scattering. We consider a thought-experiment where an incoming plane wave propagates through a chunk of random medium with typical dimension *L*. As the plane wave progresses through the heterogeneous medium, scattering operates and deflects energy in all space directions. The transmitted wavefield is distorted and attenuated.

of the Earth, the random fluctuations are superimposed on a deterministic background, usually represented by a 1-D reference model such as PREM (Dziewonski and Anderson, 1981). It is customary to assume that the spatial variations of the background velocities are slow compared to the typical scale length of the random medium. In this limit of separation of spatial scales, it is possible to consider the scattering medium as locally homogeneous. We may therefore hope to characterize large-scale variations of small-scale heterogeneities.

If we think of the medium as random and continuous, the fluctuations of a parameter such as the density ρ are characterized to lowest order by a correlation function (Rytov et al., 1989b):

$$\psi(\mathbf{x}) = \langle \rho(\mathbf{r} - \mathbf{x}/2)\rho(\mathbf{r} + \mathbf{x}/2)\rangle, \tag{1}$$

where the brackets denote an average over the ensemble of random media. In Equation 1, we have used an assumption of statistical homogeneity. In the case of a locally homogeneous medium, the correlation function may be allowed to vary slowly with the variable $\mathbf{r}$. $\psi(\boldsymbol{x})$ is maximum for $x = 0$, where it is equal to the total variance of the fluctuations denoted by ϵ^2 and goes to zero over a typical spatial scale a which is known as the correlation length. Common choices for ψ are Gaussian and exponential functions but a wide range of other possibilities exist (Klimeš, 2002). An equivalent description of the medium fluctuations is provided by the power spectrum of heterogeneities $\Phi(\mathbf{k})$ defined as:

$$\psi(\mathbf{x}) = \frac{1}{(2\pi)^3} \iiint_{-\infty}^{+\infty} \Phi(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d^3k \tag{2}$$

$\Phi(\mathbf{k})$ quantifies the distribution of the total variance of the fluctuations over all possible length scales. A very useful tool to represent the heterogeneity of fluid or solid geophysical media is the Von-Karman correlation function, whose power spectrum can be written as (Sato and Fehler, 1998):

$$\Phi(\mathbf{k}) = \frac{8\pi^{3/2}\varepsilon^2 a^3 \Gamma\left(\kappa + \frac{3}{2}\right)}{\Gamma(\kappa)(1 + k^2a^2)^{\kappa+3/2}}. \tag{3}$$

In Equation 3, ϵ denotes the RMS fluctuations, a is the correlation length and $\kappa > 0$ is the Hurst exponent which acts as a cutoff of the small-scale features in the medium. The power spectrum (Equation 3) shows a plateau at small wave numbers up to a corner value $k_c \sim 1/a$. For k much larger than k_c, the power spectrum decreases algebraically as $k^{2\kappa+3}$. Note, however, that this description does not apply to Gaussian random media. This is not a severe restriction, since Gaussian media are often regarded as too smooth to represent faithfully the heterogeneities in the Earth. This point of view is confirmed by the analysis

of well-logs (Shiomi et al., 1997), which further shows that only the parameter κ can usually be recovered from observations.

The coherent and incoherent fields

Following the statistical approach, it is natural to characterize the propagation in the random medium by calculating the moments of the wavefield u. This yields a simple classification of the various components of the random field. The first moment $\langle u \rangle$ defines the coherent wave (Rytov et al., 1989a; Sheng, 1995) and corresponds to the part of the field which survives after an average over the statistical ensemble has been performed. In a typical seismological experiment, the measured field is the sum of two terms: the ensemble average $\langle u \rangle$ and a fluctuating part u_f. The coherent field mostly contributes in a time window around the seismic primaries, which may also be termed the ballistic waves. Although the coherent field is an important component of the ballistic field, it has to be carefully distinguished from it (Derode et al., 2001). Statistical wave theory shows that the mean field decays like $e^{-L/2l}$, where L is the propagation distance and l is known as the mean free path (Sheng, 1995). The mean free path gives the typical spatial scale over which a significant amount of energy has been transferred from the coherent to the incoherent or fluctuating part of the wavefield. The incoherent part of the wavefield contributes to all portions of the seismogram and forms the coda of the earthquake records. It is important to realize that the term "incoherent" just indicates the generation of waves that propagate in a direction which differs from that of the incident wave. It does not imply that the phase information has been lost. In section Interferometry with scattered waves, we will demonstrate that the so-called incoherent field is in fact highly "coherent" in the sense that it gives rise to very useful interference effects.

The propagation regimes

The distinction between the fluctuating and ensemble average part of the wavefield is useful to characterize the propagation regime. For propagation distances typically less than the mean free path, the measured field is dominated by the coherent wave: this is the weak fluctuation regime (Rytov et al., 1989a). As shown in Figure 1, after propagating through several inhomogeneities, the most important observation is the distortion of the incident plane wave front. This gives rise to fluctuations of both phase and amplitude, from which information on the statistical properties of the medium can be extracted. For propagation distances much larger than the mean free path, the coherent wave completely vanishes and one is left with the fluctuating part only: this is the strong fluctuation regime (Rytov et al., 1989a). Since by definition the ensemble average of the fluctuating part u_f is zero, it is only through the consideration of the second moment of the field that one can model the propagation of scattering when fluctuations are strong. Specific approaches will be developed in sections Envelope modeling: Markov approximation and Envelope modeling: radiative transfer.

At this stage, it is natural to raise the following question: How can we measure the mean free path in the Earth? First, it is important to note that one has to consider at least two different mean free paths since both longitudinal and transverse waves propagate through an elastic medium. In the laboratory, there are well-calibrated techniques to measure the mean free path based on ensemble averaging. Unfortunately, these methods cannot be transposed to field experiments. Instead, the common seismological practice consists in correcting the seismograms for the travel-time fluctuations caused by long wavelength variations of the velocity before stacking or averaging the records. The field obtained after such a processing is called the traveltime-corrected mean field. Sato (1982, 1984) and Wu (1982) have demonstrated that this field decays exponentially on a scale which is much larger than the mean free path, because the traveltime correction in fact removes the long wavelength components of the medium fluctuations. A theoretical discussion of various averaging procedures and their application to seismic data is given by Shapiro and Kneib (1993).

Analysis of transmission fluctuations

As announced in introduction, we now review the various approaches that have been developed to infer the statistical properties of the Earth from an analysis of seismic wave scattering. We will first introduce a method which is well adapted to the analysis of fluctuations of amplitude and phase of teleseismic P waves recorded on a dense array. Such an approach has been pioneered in the seventies by Aki (1973) and has been subsequently developed by Wu and coworkers (Wu and Flatté, 1990; Zheng et al., 2007; Zheng and Wu, 2008). In this approach, one considers a small time window around the direct P-wave at teleseismic distance. It is usually assumed that at sufficiently large distance from the source, the wave front incident from the mantle can be approximated locally by a plane. By using sufficiently deep earthquakes, it is possible to avoid the effects of strong heterogeneities in the vicinity of the source. As the plane wave progresses through the lithosphere, the wave front is distorted and the amplitude and phase of the ballistic waves measured on the array fluctuate. Zheng and Wu (2005) have very carefully discussed the correct measurement practice. The main pulse recorded on the vertical component of the seismometer is first isolated from the rest of the signal. After Fourier transformation, the amplitude A and the phase ϕ of the pulse is obtained at a given frequency. This operation is repeated at all stations of the array, and the phase field is subsequently unwrapped spatially, i.e., possible phase discontinuities are removed. It is not correct to measure the phase of the field by picking the first arrival since, as a consequence of scattering, the medium is dispersive and the first arrivals propagate with the group velocity.

Basic observables are the fluctuations of the phase and of the logarithm of the amplitude. The reason why one considers the logarithm of the amplitude is that this quantity enters in the Rytov representation of the wavefield Rytov et al. (1989a):

$$u = u_0 e^{\psi}, \tag{4}$$

where u_0 is the reference plane wavefield in the background medium and ψ is a complex phase:

$$\mathrm{Re}\psi = \ln\frac{A}{A_0} \quad \mathrm{Im}\psi = \phi - \phi_0 \tag{5}$$

Scattering theory based on the Rytov representation is more accurate than the standard Born approximation in the transmission geometry. In Equation 5, A (resp. A_0) and ϕ (resp. ϕ_0) denote the amplitude and unwrapped phase of the field (resp. reference field). From the array measurement, one can estimate the transverse coherence function of the phase field which is defined as:

$$C_\phi(\boldsymbol{r}) = \langle \phi(\boldsymbol{x})\phi(\boldsymbol{r}+\boldsymbol{x})\rangle, \tag{6}$$

where the points $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{r}$ lie on the surface of the Earth. More generally, one can consider the transverse angular coherence function of the two phase fields corresponding to two incoming plane waves with different incident wavevectors $\boldsymbol{k}$ and $\boldsymbol{k}'$ (Wu and Flatté, 1990; Chen and Aki, 1991).

The interpretation of the phase coherence function is based on two major theoretical approaches, with different domains of validity. These approaches rely on the basic physical assumption that the medium fluctuations have a spatial scale, a, which is much larger than the probing wavelength, λ. In this regime, the scattering occurs preferentially in a small angular cone around the forward direction. As a consequence, one can usually neglect the backscattered waves and consider one-way propagation only. The crucial nondimensional parameter that governs the properties of the phase is known as the wave parameter, D, which is defined as (Rytov et al., 1989a):

$$D = \frac{\lambda L}{a^2} \tag{7}$$

In the numerator of Equation 7, the length $\sqrt{\lambda L}$ provides the typical size of the first Fresnel zone which can be interpreted as the typical volume sampled by the waves around a ray of length L. The region $D \ll 1$ can be conveniently treated with the methods of geometrical optics. In this regime, there are some well-established results for the phase coherence function of a single plane wave with vertical incidence: (1) The variance of the phase is proportional to the total variance of the medium fluctuations and increases linearly with the propagation distance L. (2) The typical correlation length of the phase is proportional to the correlation length of the random medium. This result makes sense physically, since parts of the incident wave front which are located a distance a apart – a, the correlation length– will visit different inhomogeneities. This provides a very useful method to map the typical scale of heterogeneities. The geometrical optics approach is rather flexible and allows the treatment of more complex situations. Kravtsov et al. (2003) have adapted the method to the reflexion geometry usually encountered in exploration geophysics.

The method of geometrical optics usually breaks down for $D > 1$ as caustics may develop for sufficiently large propagation distances. In order to account properly for diffraction effects, it is important to use a wave-based approach. As in the problem of transmission fluctuations in which the propagation is essentially one-way, it is convenient to consider the depth variable z as an evolution variable and to neglect the backscattered waves. In this case, the usual Helmholtz equation can be replaced by the following parabolic approximation for the wavefield:

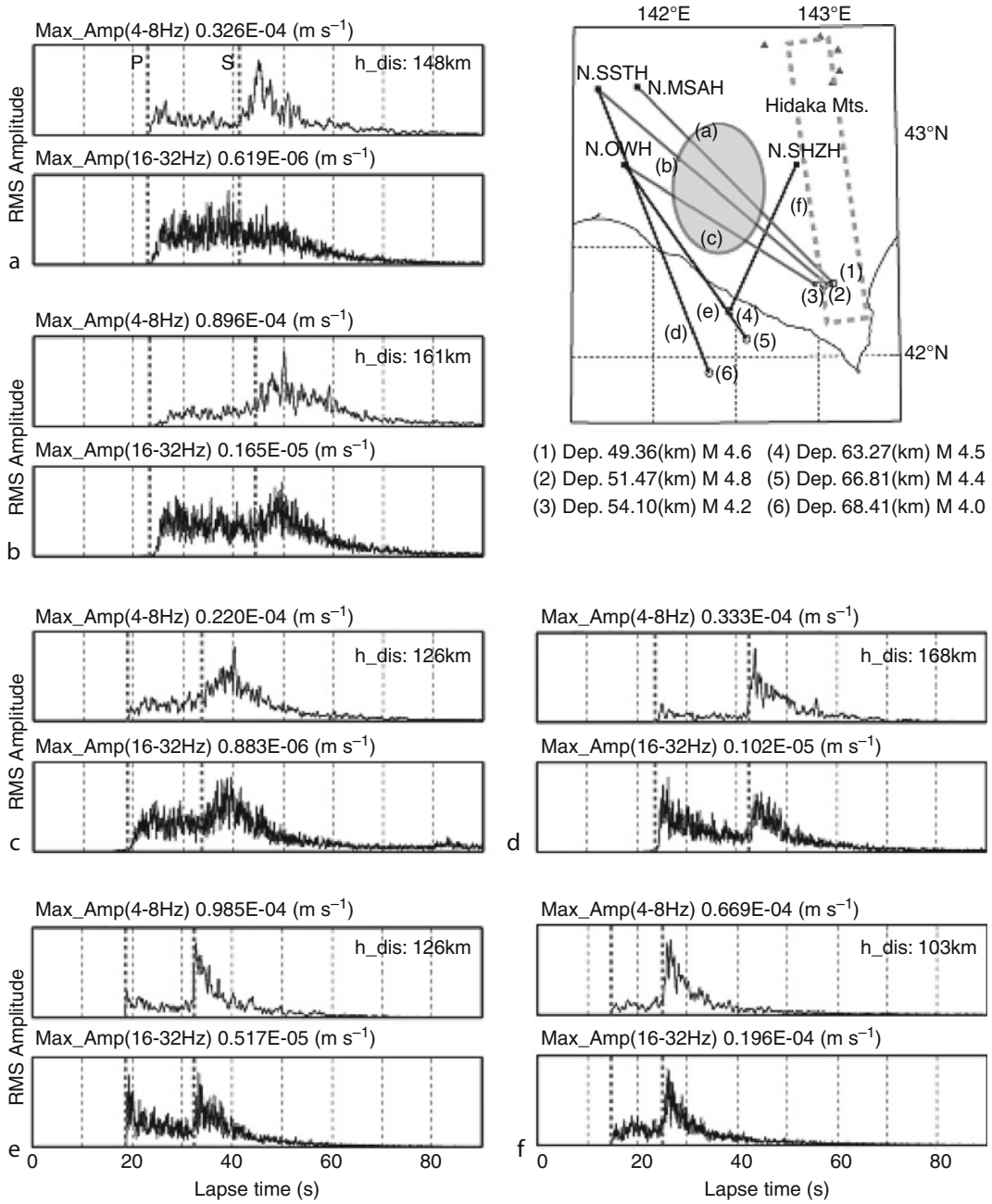
$$2ik_0\frac{\partial u}{\partial z} + \Delta_\perp u - 2k_0^2\frac{\delta c}{c_0}u = 0, \tag{8}$$

where $\Delta_\perp = \partial_x^2 + \partial_y^2$ represents the Laplace operator in the plane perpendicular to the main propagation direction. The parabolic approximation is accurate for low-angle scattering only, which is consistent with the scaling $\lambda/a \ll 1$. In this regime, a very powerful approach couples scattering theory and the Rytov representation of the wavefield to obtain expressions for the coherence function that extend the results of geometrical optics to the region $D > 1$. We refer to Wu and Flatté (1990) for further details on the derivation.

The first application to data was made by Aki who estimated the correlation length and total variance of the velocity fluctuations from the cross-coherence of amplitude and phase under the LASA array (Aki, 1973). In this pioneering work, he further approximated the correlation function of the medium with a Gaussian function. It was later realized that this choice was not always the most appropriate since many studies revealed that the more general Von-Karman correlation function could be used to describe a broad class of random media. Flatté and Wu (1988) developed a two-layer model of heterogeneity of the upper mantle based on observed transverse and angular correlation functions at NORSAR. By superimposing a layer with Hurst exponent 1/2 extending from 15 to 250 km depth upon a layer with flat power spectrum extending from 0 to 200 km depth, they were able to explain satisfactorily the joint measurements of phase and amplitude fluctuations. The most recent developments by Zheng et al. (2007) and Zheng and Wu (2008) allow the treatment of velocity gradient in the reference medium. Their general formalism offers the possibility to perform a nonparametric inversion, i.e., the full power spectrum of the fluctuations at depth can in principle be mapped from the surface measurements of angular correlation functions. For a complete description of the method as well as numerous references to the literature, we refer to the review by Zheng and Wu (2008).

Envelope modeling: Markov approximation

We now consider a regime of propagation where the fluctuating part of the field is larger than the coherent part. In this regime, an interesting phenomenon termed "envelope broadening" can be used to retrieve information on the statistical fluctuations of velocities in the medium. Enveloped broadening is illustrated in Figure 2, where small earthquakes have been recorded at epicentral distances of the order of 150 km in eastern Japan. On these data, we can observe a delay between the onset of the direct shear wave and the maximum of the amplitude. Such a delay is characteristic of the strong forward-scattering regime where the waves follow "snake-like" paths around the ballistic trajectory between source and station. As clearly shown in



Seismic Waves, Scattering, Figure 2 Example of earthquake records showing the phenomenon of envelope broadening. Paths a, b, c propagate through a very heterogeneous volcanic region. A large delay between the onset of *S* waves and the maximum amplitude is clearly observed. This phenomenon further depends on the frequency band considered. Paths d, e, f do not cross the heterogeneous region and show more impulsive envelopes with small onset to peak delay time (Reproduced from Takahashi et al. (2007)).

Figure 2, envelope broadening depends on propagation distance and frequency band. It is also noticeable that very heterogeneous regions such as volcano areas give rise to much stronger envelope broadening than the usual crust. The clear path dependence suggests the possibility to use regional variations of the onset to peak delay time to map heterogeneities in the crust, in order to develop a stochastic tomography in the strong fluctuation regime.

Such an approach has been pioneered by Sato in the late 1980s. A recent review of the topic in seismology is given by Sato and Korn (2009). Based on the parabolic approximation, Sato (1989) derived an equation for the two-point, two-frequency coherence functions of the wavefield for incident plane waves. This technique which was originally introduced in astrophysics has been termed "Markov approximation." In the case of regional propagation, it is preferable however to consider the radiation from a point source. Following Saito et al. (2002), we write the coherence function of the field as follows:

$$\Gamma(\mathbf{r}_{\perp 1},\ \mathbf{r}_{\perp 2},\ r,\ \omega_d, \omega_c) = \langle U(\mathbf{r}_{\perp 1},\ r,\ \omega_1)\langle U(\mathbf{r}_{\perp 2},\ r,\ \omega_2)^*\rangle, \tag{9}$$

where ω_c and ω_d denote the central $(\omega_1 + \omega_2)/2$ and modulation $\omega_1 - \omega_2$ frequency, respectively. The introduction of two frequencies is demanded by the nonstationarity of the envelope records. The source station distance is denoted by r and the points $\mathbf{r}_{\perp 1}$ and $\mathbf{r}_{\perp 2}$ lie in a plane perpendicular to the wave front. The time dependence of the observed coda envelopes around the frequency ω_c is obtained by setting $\mathbf{r}_{\perp 1} = \mathbf{r}_{\perp 2}$ and taking a Fourier transform over ω_d.

The basic idea of the so-called Markov approximation is to consider one-way propagation and to neglect backscattering. The different slices of the medium perpendicular to the main propagation direction independently perturb the local ray direction in a random fashion. As a consequence, the distribution of ray directions at a distance r_0 from the source depends solely on the heterogeneities that have been met for $r < r_0$ (Sato, 1989; Sato and Korn, 2009). Such an approximation is valid for strong forward-scattering, i.e., when the relation $\lambda > a$ applies. This does not mean that the Markov approximation cannot be used if the medium contains heterogeneities with scale lengths shorter than the wavelength. Indeed, the large-scale fluctuations contribute predominantly to the field measured in the transmission geometry. Scattering at large angles caused by small-scale fluctuations mainly plays the role of an apparent energy loss which adds to the intrinsic losses (Sato, 1989). In order to quantify the typical validity region of the Markov approximation, it is useful to introduce a quantity known as the transport mean free path (Sheng, 1995) and defined as:

$$l^* = \frac{l}{1 - \langle \cos\theta \rangle}, \tag{10}$$

where $\langle \cos\theta \rangle$ is the mean cosine of the scattering angle. The transport mean free path can be interpreted as the typical length beyond which the waves have lost memory of their initial propagation direction. In the case of large-scale fluctuations, the transport mean free path l^* is typically one or two orders of magnitude larger than the scattering mean free path l. For propagation distances of the order of the transport mean free path, backscattering can no longer be ignored and one must appeal to other methods to be described in section Envelope modeling: radiative transfer.

As carefully explained in the paper of Sato (1989), there are two basic phenomena that contribute to the observed envelope shape. The first one is termed "the wandering effect" and is purely statistical in nature. It corresponds to the small arrival time fluctuations in the line-of-sight propagation direction. It leads to a Gaussian spread of an initially delta-like pulse. This reshaping can only be observed upon ensemble averaging and is therefore not of primary importance in seismology. The second phenomenon is the increase of the typical duration of the seismogram with propagation distance which has been termed "pulse" – or may be more appropriately – "envelope broadening." It reflects the angular spreading of the wave propagation directions as they interact with the medium heterogeneities. For a general Von-Karman correlation function, a time scale t_m characterizes the typical duration of the seismogram (Saito et al., 2002):

$$t_m = C\left(\varepsilon^{1/(p-1)} a^{-1}\right)^{2-2/p} \omega^{-2+4/p} r_0^{1+2/p}, \tag{11}$$

where ϵ denotes the root mean square fluctuations, ω is the central frequency of the signal, r_0 is the hypocentral distance and C is a numerical pre-factor which depends on the wavespeed, and on the Hurst exponent κ. The parameter p is an increasing function of κ and varies from 1.2 to 2 for $0 < \kappa \leq 1$. Formula (11) reveals interesting characteristics of the envelope broadening phenomenon. For a smooth medium, i.e., a medium poor in small-scale features, the parameter p is close to 2 and, as a consequence, the typical envelop duration is frequency-independent. Such a result also holds for the Gaussian correlation function and is in sharp contradiction with observations. As a consequence, the frequency dependence of envelope duration offers access to the degree of roughness of Earth's heterogeneities. Accurate measurements of the exponent κ requires that absorption and – to some extent – scattering losses caused by small-scale heterogeneities be incorporated in envelope modeling (Saito et al., 2002).

In practical applications to data, a characteristic time, t_q, is measured as a function of frequency and hypocentral distance: $t_q = t_p + t_{1/2}$, where t_p is the duration from onset to peak and $t_{1/2}$ is the decay time of the envelope from the peak to RMS half-peak value. Note that the observations are usually focused on the S wave train for which pulse broadening is usually much clearer than for P waves. The observations are then compared to those numerically obtained by the Markov approximation for a wide range of correlation lengths, RMS perturbations and intrinsic quality factor. In general it is not possible to estimate

independently ϵ and a since various combinations of these two parameters may yield the same characteristic time t_m. However, by analyzing separately the distance and frequency dependence of t_q, the intrinsic quality factor Q_s, the parameter p, and in turn, the Hurst exponent κ can be estimated. For applications to seismic observations, we refer to, e.g., Scherbaum and Sato (1991); Obara and Sato (1995); Saito et al. (2005).

As illustrated in Figure 2, for comparable hypocentral distances and frequency bands, the envelope shapes may strongly depend on the direct ray path. Such an observation calls for the introduction of large-scale variations of small-scale heterogeneities. Such an approach has been taken by Takahashi et al. (2007) who mapped strongly heterogeneous regions beneath quaternary volcanoes in northeastern Japan from the observed path dependence of the characteristic broadening time t_q. A complete tomographic approach to estimate the power spectrum of small-scale heterogeneities from envelope broadening data was further developed and applied to northeastern Japan by Takahashi et al. (2009).

Gusev and Abubakirov (1999a, b) developed a tomographic method to map the transport mean free path in the Earth based on body wave envelope broadening. Their approach is simpler since they consider the inversion of a single parameter. From the knowledge of the frequency dependence of the the transport mean free path, it is possible to put constraints on the parameters ϵ, a, and κ. The Markov method has recently been extended to the case of vector waves by Sato (2006, 2007, 2008). The theoretical approach relies on the introduction of scalar and vector potentials for P and S waves, respectively. The employed Helmholtz decomposition neglects the conversions between P and S waves, which is justified in the high-frequency regime $\lambda \gg a$, up to a characteristic coupling time τ^{PS} between P and S waves. Extensions of the Markov method to the vectorial case allow the modeling of the observed energy envelopes of the transverse components of the wavefield after the onset of the direct P-wave. The observation of such transverse energy is another manifestation of the randomization of the ray directions in a scattering medium. Alternative explanations for this phenomenon are mode conversions or anisotropy. In the case of a statistically homogeneous Gaussian medium with thickness h, Sato (2006) showed that the ratio between the peak total intensity and the peak intensity on the transverse component of the P pulse scales like $\epsilon^2 h/a$. Interestingly, this relation is free from absorption, since it cancels out in the intensity ratio. Based on this approach Kubanza et al. (2007) made a global study of energy envelopes of transverse components in the P wave train observed at teleseismic distances and found good correlation between the strength of heterogeneity and the tectonic setting.

Envelope modeling: radiative transfer

When the propagation time and distances become very large compared to the ballistic time, or when the acquisition geometry is in the reflection mode, it becomes crucial to take into account scattering at large angle. A powerful approach to model seismogram envelopes in scattering media in such cases is provided by radiative transfer theory. In seismology, radiative transfer was introduced in the mid-1980s by Wu (1985) with a first application to data by Wu and Aki (1988). Physically, the equation of transfer is a local statement of energy conservation in a random medium and incorporates an arbitrarily high number of scattering events. The radiative transfer equation also takes into account the anisotropy of the scattering process, in particular large-angle scattering. Therefore, it enables the complete modeling of seismogram envelopes including coda waves, i.e., the late arrivals of the seismogram. A review of the radiative transfer approach in seismology is given by Margerin (2005).

The study of coda waves has been pioneered by Aki (1969) and Aki and Chouet (1975). In this last reference, the authors developed two famous models of scattering: the single-scattering approximation and the diffusion approximation. The single-scattering model applies for propagation distances of the order of, or less than the mean free path. This model has been very popular until studies by Gusev and Abubakirov (1987) and Hoshiba (1991) pointed out the importance of multiple scattering. Recently, the diffusion model has known a revived interest in connection with seismic experiments on volcanoes (Wegler and Lühr, 2001; Wegler, 2004). The single-scattering and diffusion models are respectively short-time and large-time asymptotics of the radiative transfer equation. Mathematically, radiative transfer takes the form of an integro-differential equation for an angularly resolved energy flux known as the specific intensity $I(\mathbf{r}, \hat{\mathbf{k}}, t, \omega)$. The specific intensity quantifies the amount of energy flowing around the direction $\hat{\mathbf{k}}$ in a small frequency band $[\omega, \omega + \delta\omega]$ during a time interval $[t, t + \mathrm{d}t]$ through a small surface element located at $\mathbf{r}$. Although the concept of specific intensity was first introduced on a phenomenological basis, it is now well understood that it is in fact connected to the Wigner-Ville distribution of the wavefield (Ryzhik et al., 1996):

$$I(\mathbf{r},\hat{\mathbf{k}},t,\omega) = C \iint \langle u(t-\tau/2,\mathbf{r}+\mathbf{x}/2)u(t+\tau/2,\mathbf{r}-\mathbf{x}/2)^{*}\rangle e^{-i\mathbf{k}\cdot\mathbf{x}+i\omega\tau} d^3x d\tau, \tag{12}$$

where the brackets denote an ensemble average. The key to understand this definition is the notion of separation of scales. In Equation 12, one must think of the intensity as a slowly varying function of the time t, as compared to the fast oscillations of the wavefield described by the frequency ω. Similarly, the average intensity varies slowly in space as compared to the wavelength. Such a separation of scales is well satisfied by high-frequency seismograms and guarantees that the Wigner distribution can be interpreted as a local power spectrum of the wavefield.

It is important to note that interference effects are neglected in the transfer approach. This can be justified on the basis of the small phase shifts that occur upon scattering and which will be different for waves visiting different heterogeneities.

An important step in the development of radiative transfer was the introduction of the Monte Carlo method, which is a flexible numerical technique to solve the transfer equation. It has been introduced in seismology by Gusev and Abubakirov (1987). The Monte Carlo method simulates the random walk of particles, which change direction of propagation at the end of each step. The step length has an exponential distribution with parameter the scattering mean free path, and the change of direction upon scattering is provided by a probabilistic interpretation of the differential scattering cross-section. The latter quantity can be obtained from the Born approximation at frequencies below the geometrical optics limit. The validity of this approximation has been discussed by Wegler et al. (2006a) and Przybilla and Korn (2008). The Monte Carlo method can be shown to yield an exact solution of the transfer equation (Papanicolaou et al., 2000). The variance of the result typically decreases like the square root of the number of simulated random walks.

An important application of the radiative transfer approach was developed by Fehler et al. (1992) and Hoshiba (1993) who proposed a method to infer the mean free path and intrinsic quality factor in the crust from the space time dependence of the energy radiated by small earthquakes. The method is based on the estimation of the energy in three successive time windows starting with the direct *S* wave arrivals. Such multiple lapse-time window analysis has been applied in various regions of the world to give estimates of the level of heterogeneity in the crust. Typically, the mean free path of the crust around 1 Hz ranges from a few tens to a few hundred kilometers. The method has some limitations since it assumes isotropic scattering and a uniform distribution of scatterers in a half-space.

It was pointed out by Gusev and Abubakirov (1987), Abubakirov and Gusev (1990) and Hoshiba (1995) that the observed shape of envelopes of body waves is incompatible with isotropic scattering, since no broadening occurs in this case. Based on the comparison between modeled and observed *S* wave seismogram envelopes, Gusev and Abubakirov (1996) were able to infer a Hurst exponent ranging between 1/4 and 1/2 in the crust. In the case of crustal propagation at regional distances, the field is dominated by guided waves. To take into account the role of stratification and scattering in the Earth, Hoshiba (1997), Margerin et al. (1998), and Yoshimoto (2000) have shown how to incorporate depth-dependent velocity and mean free path in the Monte Carlo method. Margerin et al. (1999) numerically solved the radiative transfer equation in the crustal geometry and confirmed the importance of the leakage of energy at the Moho to explain the decay of the coda, as first proposed by Korn (1990).

Other applications of radiative transfer pertain to the study of the seismic source. Nakahara et al. (1998) proposed a method based on radiative transfer theory to infer the distribution of high-frequency radiation on the fault plane of large earthquakes. They considered the different part of the fault to be independent and modeled the propagation using a multiple isotropic scattering model. Sens-Schönfelder and Wegler (2006a) devised a method to infer the seismic moment of small crustal earthquakes recorded at regional distance. Another important application concerns the modeling of wave propagation in very heterogeneous structures such as volcanoes, where the mean free path can become as small as a few hundred meters. A review of recent development in the field is provided by Del Pezzo (2008). It is to be noted that in cases where the typical propagation distance is much larger than the mean free path, the radiative transfer equation can be simplified to a simple scalar diffusion equation for the total energy density of *P* and *S* waves. The diffusion approach to the modeling of seismic energy propagation in volcanoes has been advocated by Wegler (2004).

Radiative transfer theory is still in an active stage of development. A radiative transfer equation for elastic waves has been introduced in seismology by Zeng (1993); Sato (1994); and Ryzhik et al. (1996). This equation takes into account in a rigorous way the coupling between *P* and *S* waves. A Monte Carlo method to solve this elastic transfer equation was developed by Margerin et al. (2000). A comparison of finite-difference and Monte Carlo simulations by Przybilla and Korn (2008) demonstrated the accuracy of the radiative transfer approach to simulate the coupling and multiple scattering of *P* and *S* waves. An extension of the radiative transfer equation to incorporate the coupling between surface and body waves is a major challenge. An important step in this direction was made by Maeda et al. (2007) who incorporated the coupling between Rayleigh and body waves in the single-scattering approximation. The impact of anisotropic scale lengths has been studied by Margerin (2006) who derived a radiative transfer equation with an angular-dependent scattering mean free path.

Global-scale scattering

The subject of seismic scattering at the global scale was first developed in the early 1970s in connection with short-period precursors to phases such as *PP* (King et al., 1975) and *PKP* (Cleary and Haddon, 1972). Accompanying the improvement of the global seismic network and the availability of high-quality data, evidences for the importance of scattering in the deep Earth have accumulated in recent years. The study of precursors to deep phases such as *PKP* (Hedlin et al., 1997) and *PKKP* (Earle and Shearer, 1997) has been used to put constraints on the heterogeneities in the deep mantle and core. Observation and modeling of anomalous propagation of seismic waves in the subducting lithosphere have put forward the role of scattering by elongated heterogeneities in slabs

(Furumura and Kennett, 2005). The presence of strong scatterers in Earth's inner core has been revealed by a number of studies (Vidale and Earle, 2000; Koper et al., 2004; Poupinet and Kennett, 2004). Vidale and Earle (2000) used data from nuclear explosions recorded on the LASA array to demonstrate that the long-lasting signal following the ballistic *PKiKP* phase was not caused by the response of the crust but were genuine scattered waves radiated from the inner core. They showed examples of slowly-emergent and long-lasting *PKiKP* signals recorded at an epicentral distance of about 60°. This last point is particularly important, since it corresponds to incidence angles of *P* waves at the inner core boundary such that the reflection coefficient vanishes. The coda of *PKiKP* thus provides a direct proof of the presence of scatterers in the inner core.

Since the deep Earth interior is usually thought to be much less heterogeneous than the lithosphere, much of the modeling work has been focused on the single-scattering approximation (Hedlin et al. (1997); Vidale and Earle (2000); Leyton and Koper (2007)). The radiative transfer approach was introduced in global seismology by Margerin and Nolet (2003a, b) to improve the modeling of the precursors to the *PKP* wave. In particular, they demonstrated the ability of radiative transfer to incorporate complicated ray geometry such as the one imposed by the severe drop of velocity at the core–mantle boundary. The Monte Carlo method was further developed by Shearer and Earle (2004) who proposed a scheme to simulate the multiple scattering of elastic waves at the global scale, including the coupling between *P* and *S* waves. Their method was applied to the study of the coda envelopes of teleseismic *P* waves. To explain the observations, they proposed a 1-D model of Earth small-scale heterogeneity composed of a heterogeneous 200 km top-layer (4% RMS perturbations) overlying a less heterogeneous 400-km layer (3% RMS perturbations). At the base of the mantle, a 600 km thick layer with 0.5% RMS perturbations is required to achieve a good fit to the data. Shearer and Earle (2004) concluded that attenuation is dominated by scattering in the upper-mantle and by absorption in the lower mantle.

In addition to global studies of scattering of short-period waves, an interest in the scattering of long-period surface waves has recently arisen. Maeda et al. (2006) performed an (*f,k*) analysis of the vertical components of three large earthquakes recorded on the Japanese Hi-Net in the 90–180 s period band. They find that the signals recorded between the first multi-orbiting Rayleigh wave arrivals are composed of waves coming from all directions with phase velocities of the order of the Rayleigh wave fundamental mode. This suggests that in the period band of interest, scattered Rayleigh waves are present in the data at short lapse time. After typically 30,000 s, the multi-orbiting Rayleigh waves have strongly decayed. The (*f,k*) analysis demonstrates that the signal is dominated by waves with very large phase velocities, thereby implying the dominance of higher-mode Rayleigh waves. Such a phenomenon is to be expected on the ground that higher modes sample the deeper parts of the Earth with high intrinsic quality factor. To model the coda of multi-orbiting Rayleigh waves, Sato and Nishino (2002) developed a radiative transfer equation on the sphere assuming the multiple isotropic scattering of the fundamental mode Rayleigh wave. In their approach, the coupling with higher modes is neglected. They observe a systematic discrepancy between observed and modeled seismogram envelopes which also suggests that higher modes dominate the data at lapse times larger than 30,000 s.

Interferometry with scattered waves

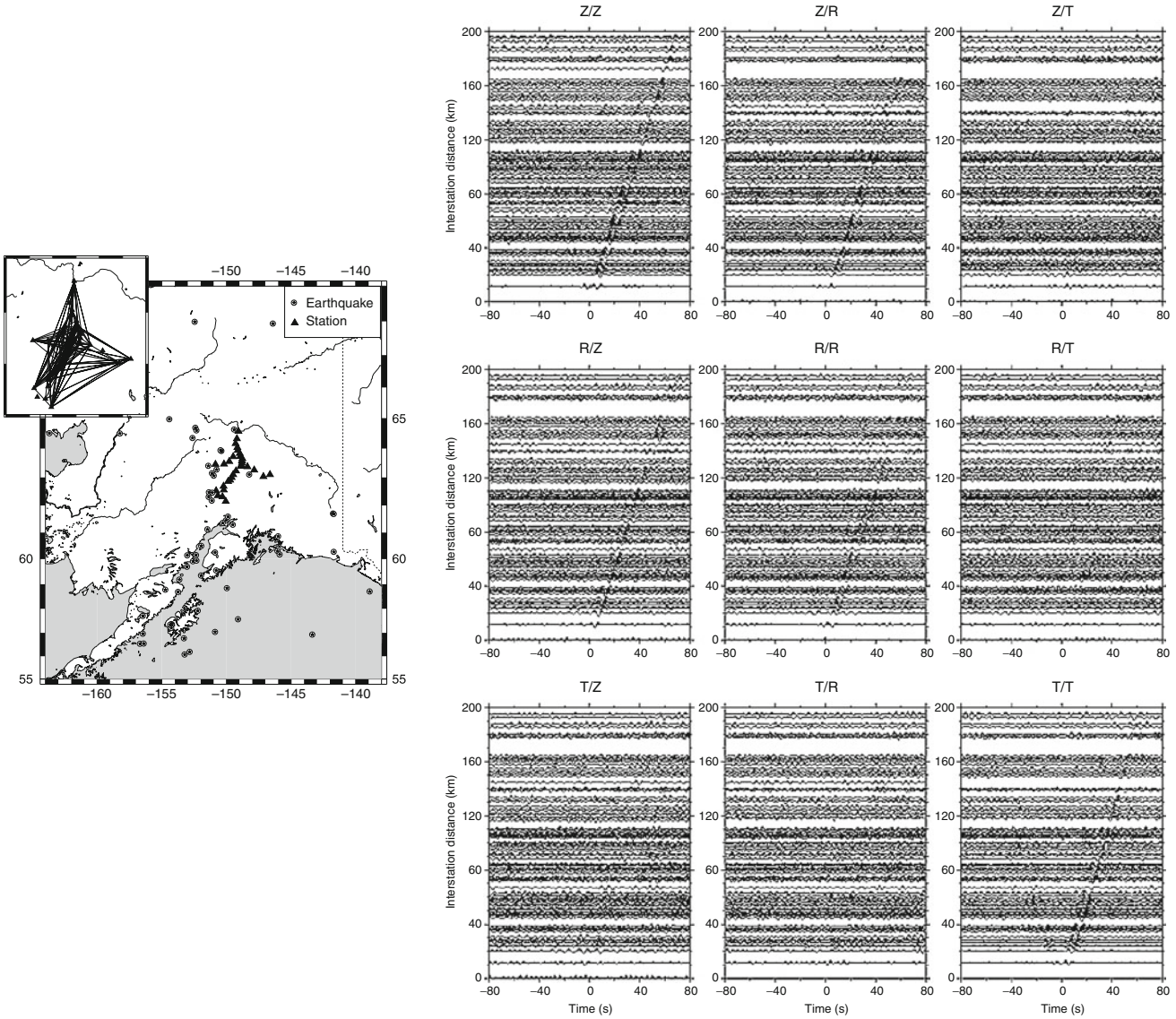
Green function retrieval

As explained in section Introduction, the incoherent field refers to the waves that average out when a mean over an ensemble of random media is performed. These waves are incoherent in the sense that, at a given point of the medium, they propagate in space directions that differ from that of the mean field. But "incoherence" does not imply that the phase information has been lost (Campillo, 2006). Campillo and Paul (2003) were the first to demonstrate the existence of correlations between the coda waves recorded at two distant points in the heterogeneous crust. Such correlations emerge after sufficient temporal and source averaging. Following Lobkis and Weaver (2001), Campillo and Paul (2003) proposed that the correlation tensor of the coda wavefields recorded at two points *A* and *B* is in fact proportional to the Green's tensor G_{ij} between these two points. In the frequency domain, such a relation can be mathematically formulated as follows:

$$\left\langle u_i(B,\omega)u_j(A,\omega)^*\right\rangle \propto \mathrm{Im}G_{ij}(B,A,\omega). \tag{13}$$

The proportionality factor depends on the details of the source spectrum. The occurrence of the imaginary part of the Green's function on the right-hand side of Equation 13 physically means that the correlation of two wavefields is proportional to a combination of the retarded and advanced Green's functions. The appearance of the advanced Green's function is most easily understood by noting the equivalence between correlations and time-reversal experiments. We will not discuss this analogy, but the interested reader can refer to Derode et al. (2003).

In seismology, the brackets in Equation 13 usually denote an average over time windows in the coda and over earthquake sources. An example of emergence of the retarded and advanced Green's function in the correlation of coda waves recorded on a temporary network deployed in Alaska is illustrated in Figure 3. The maximum of the correlation is always shown for positive times, but it is nevertheless possible to distinguish some coherent arrivals at negative times. The reconstruction of the Green's function from coda waves is intimately related to the concept of equipartition, which is a fundamental consequence of multiple scattering. Loosely stated, equipartition stipulates that plane *P, S*, Rayleigh, and Love



Seismic Waves, Scattering, Figure 3 Example of Green's function reconstruction from coda waves recorded in Alaska. A location map of the experiment is shown on the *left*. The inset shows all the station pairs analyzed in this study. On the *right*, all possible terms of the correlation tensor are represented as a function of the correlation time on the horizontal axis and the epicentral distance between the two stations on the vertical axis. For each station pair, the three components of the seismometers have been rotated onto the radial (R), transverse (T), and vertical (Z) reference frame (Reproduced from Paul et al. (2005)).

waves coming from all possible directions compose the wavefield. That the Green's function of the medium is recovered if the equipartition state is reached has been shown for an infinite elastic medium by Sánchez-Sesma and Campillo (2006). Based on the single-scattering approximation for scalar waves, Sato (2009) demonstrated that the correlation of coda waves recorded at two stations enables the reconstruction of the ballistic Green's function between the two stations. Snieder (2004b) derived the reconstruction of the Green's function from coda waves based on the stationary phase approximation. An experimental verification of the equipartition principle has been performed by Hennino et al. (2001). In Figure 3, we can observe a clear time asymmetry in the reconstructed Green's function. This seems to contradict the equipartition principle. Such a time asymmetry has been studied by Paul et al. (2005). These authors demonstrate that a flux of energy directed from the earthquake source to the station can persist even at very large lapse time in the coda. Such an energy flux breaks the temporal symmetry of the correlations unless the distribution of earthquakes around the station covers all azimuths.

Monitoring temporal variations

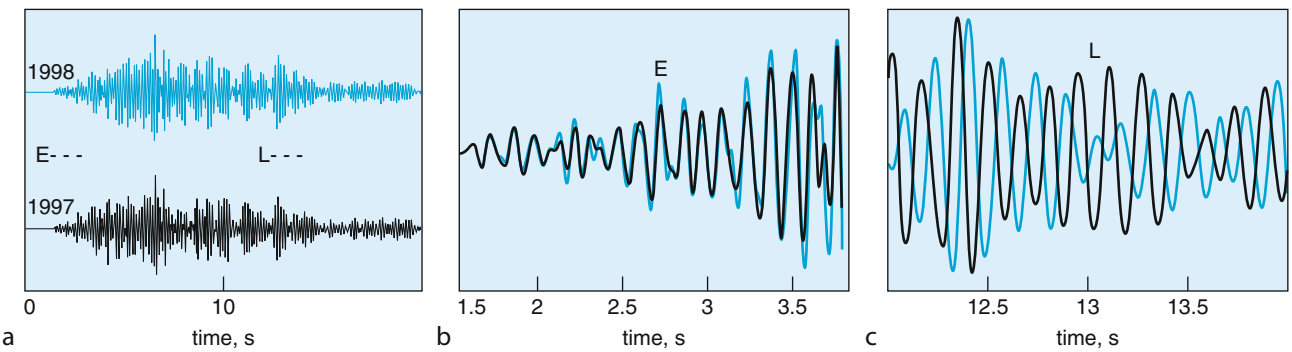
An interesting application of wave multiple scattering pertains to the detection of weak changes in dynamic media, i.e., media evolving with time. The methods which make use of scattered waves to probe changes in complex media have been termed "diffusing acoustic wave spectroscopy" in acoustics and "coda wave interferometry" (CWI) in seismology. In this article, we focus on applications to seismic waves. A theoretical approach of the method is provided by Snieder (2006). For a very accessible review of this broad topic in physics, we refer to Snieder and Page (2007). A comprehensive review of monitoring applications in seismology is given by Poupinet et al. (2008).

The key idea of CWI is to take advantage of the long propagation paths of scattered waves, which accumulate small phase differences on their way from source to station. An example of monitoring the medium changes at the Merapi volcano is shown in Figure 4. In this experiment, a repeatable source is used to generate seismic waves that propagate through the volcano. The fact that the source is highly reproducible is important, as any small difference between seismograms recorded at different time instants can be interpreted as a change of the propagating medium. Natural repeating sources such as earthquake doublets can also be exploited to probe medium changes. The doublets correspond to small earthquakes breaking the same fault patch at different times of occurrence (Poupinet et al., 1984). As shown in Figure 4, the changes are virtually undetectable on the first arrivals. In the coda, the medium changes are manifested by a small time delay between two very similar waveforms. In the case of the Merapi data shown in Figure 4, an analysis of the delay times as a function of the time lag in the coda reveals that the two seismograms are nearly stretched versions of one another. This means that the two waveforms can be related by a similarity transformation $t \rightarrow t(1+\epsilon)$ of the time variable. Such an observation can be interpreted as a very small change of the wave velocity δv in the propagating medium. The relation between the stretching parameter ϵ and the velocity variation reads (Sens-Schönfelder and Wegler 2006b):

$$\epsilon = -\frac{\delta v}{v} \tag{14}$$

In practice, the velocity change can be estimated by measuring the delay time which maximizes the correlation between the two signals. Such a procedure can be applied to moving time windows in the coda, of typical duration, a few central periods. The relation (14) is valid for acoustic waves only. A formula valid for elastic coda waves in the equipartition regime has been derived by Snieder (2002). Ratdomopurbo and Poupinet (1995), and Wegler, et al. (2006b) found that the shear velocity increases prior to eruptions at the Merapi volcano, an observation interpreted as a pressurization effect due to the magma ascent. Typically, relative velocity changes of the order of 10^{-5}–10^{-4} can be detected (Poupinet et al., 2008). This makes coda wave interferometry the ideal tool to study the rheology of materials (Snieder et al., 2002).

As a final example of the use of scattered waves, we consider a method which has been termed "passive image interferometry" (Sens-Schönfelder and Wegler, 2006b) which combines coda wave interferometry with the principle of Green's function reconstruction from coda waves. Passive interferometry uses ambient noise – instead of coda waves – to reconstruct the Green's function. The coda part of the reconstructed Green's function is subsequently used to monitor temporal variations in the medium. A great advantage of this method as compared to the usual coda wave interferometry lies in the fact that noise records allow a continuous reconstruction of the Green's function over very long time periods. Brenguier et al. (2008) demonstrated the possibility to detect very



Seismic Waves, Scattering, Figure 4 Illustration of the principle of coda wave interferometry. A controlled and repeatable source is used to probe the Merapi volcano with seismic waves. The signals recorded at a time interval of 1 year are shown by *blue and black lines*. The insets show the comparisons of waveforms in the early part of the signal (E) and in the late part of the signal (L). In the later portion, a clear time shift between the two traces can be observed (Reproduced from Snieder (2004a)).

small medium changes at Le Piton de la Fournaise volcano in la Réunion Island. Applications to the study of the recovery phenomena in sediments affected by strong ground motion are developed in Sawazaki et al. (2009).

Summary

This article presents a nontechnical introduction to the various phenomena and methods pertaining to the scattering of seismic waves. The basic concepts of seismic scattering are introduced with the aid of a thought-experiment. The topics covered include the interpretation of transmission fluctuations, the broadening of envelope at regional distances, the interpretation of coda waves at local and global scales, and the use of scattered waves in seismic interferometry. The methods developed in seismic scattering offer powerful tools for the stochastic imaging of Earth's structure at spatial scales that cannot be resolved by traditional tomographic methods. In addition, scattered seismic waves are extremely sensitive to slight temporal changes of the medium. Using the interferometric approach, multiply-scattered coda waves can be used to monitor the evolution of the dynamic Earth.

Bibliography

Abubakirov, I., and Gusev, A., 1990. Estimation of scattering properties of lithosphere of Kamchatka based on Monte-Carlo simulation of record envelope of a near earthquake. *Physics of the Earth and Planetary Interiors*, **64**, 52–67.

Aki, K., 1969. Analysis of the seismic coda of local earthquakes as scattered waves. *Journal of Geophysical Research*, **74**, 615–618.

Aki, K., 1973. Scattering of P waves under the Montana LASA. *Journal of Geophysical Research*, **78**, 1334–1346.

Aki, K., and Chouet, B., 1975. Origin of coda waves, sources and attenuation. *Journal of Geophysical Research*, **80**, 3322–3342.

Brenguier, F., Shapiro, N. M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O., and Nercessian, A., 2008. Toward forecasting volcanic eruptions using seismic noise. *Nature Geoscience*, **1**, 126–130.

Campillo, M., 2006. Phase and correlation in 'Random' seismic fields and the reconstruction of the green function. *Pure and Applied Geophysics*, **163**, 475–502.

Campillo, M., and Paul, A., 2003. Long range correlations in the diffuse seismic coda. *Science*, **299**, 547–549.

Chen, X., and Aki, K., 1991. General coherence functions for amplitude and phase fluctuations in a randomly heterogeneous medium. *Geophysical Journal International*, **105**, 155–162.

Cleary, J., and Haddon, R., 1972. Seismic wave scattering near the Core-Mantle Boundary: a new interpretation of precursors to PKP. *Nature*, **240**, 549–551.

Del Pezzo, E., 2008. Seismic wave scattering in volcanoes. In Sato, H., Fehler, M., and Dmowska, R. (eds.), *Earth Heterogeneity and Scattering Effect on Seismic Waves*. New York: Academic. Advances in Geophysics, Vol. 50, pp. 353–369.

Derode, A., Tourin, A., and Fink, M., 2001. Random multiple scattering of ultrasound.I. Coherent and ballistic waves. *Physical Review E*, **64**, 36–605.

Derode, A., Larose, E., Tanter, M., de Rosny, J., Tourin, A., Campillo, M., and Fink, M., 2003. Recovering the Green's function from field-field correlations in an open scattering medium. *The Journal of the Acoustical Society of America*, **113**, 2973–2976.

Dziewonski, A., and Anderson, D., 1981. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, **25**, 297–356.

Earle, P., and Shearer, P., 1997. Observations of PKKP precursors used to estimate small-scale topography on the core-mantle boundary. *Science*, **277**, 667–670.

Fehler, M., Hoshiba, M., Sato, H., and Obara, K., 1992. Separation of scattering and intrinsic attenuation for the Kanto-Tokai region, Japan, using measurements of S-wave energy versus hypocentral distance. *Geophysical Journal International*, **108**, 787–800.

Flatté, S., and Wu, R., 1988. Small-scale structure in the lithosphere and asthenosphere deduced from arrival time and amplitude fluctuations at NOR-SAR. *Journal of Geophysical Research*, **93**, 6601–6614.

Furumura, T., and Kennett, B., 2005. Subduction zone guided waves and the heterogeneity structure of the subducted plate: Intensity anomalies in northern Japan. *Journal of Geophysical Research*, **110**, B10302.

Gusev, A., and Abubakirov, I., 1987. Monte-Carlo simulation of record envelope of a near earthquake. *Physics of the Earth and Planetary Interiors*, **49**, 30–36.

Gusev, A., and Abubakirov, I., 1996. Simulated envelopes of non-isotropically scattered body waves as compared to observed ones: another manifestation of fractal heterogeneity. *Geophysical Journal International*, **127**, 49–60.

Gusev, A., and Abubakirov, I., 1999a. Vertical profile of effective turbidity reconstructed from broadening of incoherent body-wave pulses. I- General approach and the inversion procedure. *Geophysical Journal International*, **136**, 295–308.

Gusev, A., and Abubakirov, I., 1999b. Vertical profile of effective turbidity reconstructed from broadening of incoherent body-wave pulses. II- Application to Kamchatka data. *Geophysical Journal International*, **136**, 309–323.

Hedlin, M., Shearer, P., and Earle, P., 1997. Seismic evidence for small-scale heterogeneity throughout the Earth's mantle. *Nature*, **387**, 145–150.

Hennino, R., Trégourès, N., Shapiro, N. M., Margerin, L., Campillo, M., van Tiggelen, B. A., and Weaver, R. L., 2001. Observation of equipartition of seismic waves. *Physical Review Letters*, **86**, 3447–3450.

Hoshiba, M., 1991. Simulation of multiple scattered coda wave excitation based on the energy conservation law. *Physics of the Earth and Planetary Interiors*, **67**, 123–136.

Hoshiba, M., 1993. Separation of scattering attenuation and intrinsic absorption in Japan using the multiple lapse time window analysis of full seismogram envelope. *Journal of Geophysical Research*, **98**, 15809–15824.

Hoshiba, M., 1995. Estimation of nonisotropic scattering in western Japan using coda wave envelopes: application of a multiple nonisotropic scattering model. *Journal of Geophysical Research*, **100**, 645–657.

Hoshiba, M., 1997. Seismic coda wave envelope in depth-dependent S wave velocity structure. *Physics of the Earth and Planetary Interiors*, **104**, 15–22.

King, D., Haddon, R., and Husebye, E., 1975. Precursors to PP. *Physics of the Earth and Planetary Interiors*, **10**, 103–127.

Klimeš, L., 2002. Correlation functions of random media. *Pure and Applied Geophysics*, **159**, 1811–1831.

Koper, K., Franks, J., and Dombrovskaya, M., 2004. Evidence for small-scale heterogeneity in Earth's inner core from a global study of PKiKP coda waves. *Earth and Planetary Science Letters*, **228**, 227–241.

Korn, M., 1990. A modified energy flux model for lithospheric scattering of teleseismic body waves. *Geophysical Journal International*, **102**, 165–175.

Kravtsov, Y., Müller, T., Shapiro, S., and Buske, S., 2003. Statistical properties of reflection traveltimes in 3-D randomly inhomogeneous and anisomeric media. *Geophysical Journal International*, **154**, 841–851.

Kubanza, M., Nishimura, T., and Sato, H., 2007. Evaluation of strength of heterogeneity in the lithosphere from peak amplitude analyses of teleseismic short-period vector P waves. *Geophysical Journal International*, **171**, 390–398.

Leyton, F., and Koper, K., 2007. Using PKiKP coda to determine inner core structure: 2. Determination of QC. *Journal of Geophysical Research*, **112**, B05317.

Lobkis, O. I., and Weaver, R. L., 2001. On the emergence of the Green's function in the correlations of a diffuse field. *The Journal of the Acoustical Society of America*, **110**, 3011–3017.

Maeda, T., Sato, H., and Ohtake, M., 2006. Constituents of vertical-component coda waves at long periods. *Pure and Applied Geophysics*, **163**, 549–566.

Maeda, T., Sato, H., and Nishimura, T., 2007. Synthesis of coda wave envelopes in randomly inhomogeneous elastic media in a half-space: single scattering model including Rayleigh waves. *Geophysical Journal International*, **172**, 130–154.

Margerin, L., 2005. Introduction to radiative transfer of seismic waves. In Nolet, G., and Levander, A. (eds.), *Seismic Earth: Analysis of Broadband Seismograms*. Washington: American Geophysical Union. Geophysical Monograph, Vol. 157, pp. 229–252.

Margerin, L., 2006. Attenuation, transport and diffusion of scalar waves in textured random media. *Tectonophysics*, **416**, 229–244.

Margerin, L., and Nolet, G., 2003a. Multiple scattering of high-frequency seismic waves in the deep earth: modeling and numerical examples. *Journal of Geophysical Research*, **108**, 2234.

Margerin, L., and Nolet, G., 2003b. Multiple scattering of high-frequency seismic waves in the deep earth: PKP precursor analysis and inversion for mantle granularity. *Journal of Geophysical Research*, **108**, 2514.

Margerin, L., Campillo, M., and van Tiggelen, B. A., 1998. Radiative tranfer and diffusion of waves in a layered medium: new insight into coda Q. *Geophysical Journal International*, **134**, 596–612.

Margerin, L., Campillo, M., Shapiro, N., and van Tiggelen, B. A., 1999. Residence time of diffuse waves in the crust as a physical interpretation of coda Q: application to seismograms recorded in mexico. *Geophysical Journal International*, **138**, 343–352.

Margerin, L., Campillo, M., and van Tiggelen, B. A., 2000. Monte Carlo simulation of multiple scatterring of elastic waves. *Journal of Geophysical Research*, **105**, 7873–7892.

Nakahara, H., Nishimura, T., Sato, H., and Ohtake, M., 1998. Seismogram envelope inversion for the spatial distribution of high-frequency energy radiation from the earthquake fault: application to the 1994 far east off Sanriku earthquake, Japan. *Journal of Geophysical Research*, **103**, 855–867.

Obara, K., and Sato, H., 1995. Regional differences of random inhomogeneities around the volcanic front in the Kanto-Tokai area, Japan, revealed from the broadening of *S* wave seismogram envelopes. *Journal of Geophysical Research*, **100**, 2103–2121.

Papanicolaou, G., Bal, G., and Ryzhik, L., 2000. Probabilistic theory of transport processes with polarization. *SIAM Journal on Applied Mathematics*, **60**, 1639–1666.

Paul, A., Campillo, M., Margerin, L., Larose, E., and Derode, A., 2005. Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves. *Journal of Geophysical Research*, **110**, B08302.

Poupinet, G., and Kennett, B., 2004. On the observation of high frequency PKiKP and its coda in Australia. *Physics of the Earth and Planetary Interiors*, **146**, 497–511.

Poupinet, G., Ellsworth, W., and Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California. *Journal of Geophysical Research*, **89**, 5719–5731.

Poupinet, G., Got, J.-L., and Brenguier, F., 2008. Monitoring temporal variations of physical properties in the crust by cross-correlating the waveforms of seismic doublets. In Sato, H., Fehler, M. and Dmowska, R. (eds.), *Advances in Geophysics : Earth Heterogeneity and Scattering Effect on Seismic Waves*. New York: Academic Press, Vol. 50, pp. 374–399.

Przybilla, J., and Korn, M., 2008. Monte Carlo simulation of radiative energy transfer in continuous elastic random media-three-component envelopes and numerical validation. *Geophysical Journal International*, **173**, 566–576.

Ratdomopurbo, A., and Poupinet, G., 1995. Monitoring a temporal change of seismic velocity in a volcano: application to the 1992 eruption of Mt. Merapi (Indonesia). *Geophysical Research Letters*, **22**, 775–778.

Rytov, S. M., Kravtsov, Y. A., and Tatarskii, V. I., 1989a. *Principles of Statistical Radiophysics*. New York: Springer. Wave Propagation through Random Media, Vol. 4.

Rytov, S. M., Kravtsov, Y. A., and Tatarskii, V. I., 1989b. *Principles of Statistical Radiophysics*. New York: Springer. Elements of Random Fields, Vol. 1.

Ryzhik, L., Papanicolaou, G., and Keller, J. B., 1996. Transport equation for elastic and other waves in random media. *Wave Motion*, **24**, 327–370.

Saito, T., Sato, H., and Ohtake, M., 2002. Envelope broadening of spherically outgoing waves in three-dimensional random media having power law spectra. *Journal of Geophysical Research*, **107**, 2089.

Saito, T., Sato, H., Ohtake, M., and Obara, K., 2005. Unified explanation of envelope broadening and maximum-amplitude decay of high-frequency seismograms based on the envelope simulation using the Markov approximation: forearc side of the volcanic front in northeastern Honshu, Japan. *Journal of Geophysical Research*, **110**, B01304.

Sánchez-Sesma, F. J., and Campillo, M., 2006. Retrieval of the green function from cross correlation: the canonical elastic problem. *Bulletin Seismological Society of America*, **96**, 1182–1191.

Sato, H., 1982. Amplitude attenuation of impulsive waves in random media based on travel time corrected mean wave formalism. *The Journal of the Acoustical Society of America*, **71**, 559–564.

Sato, H., 1984. Attenuation and envelope formation of three-component seismograms of small local earthquakes in randomly inhomogeneous lithosphere. *Journal of Geophysical Research*, **89**, 1221–1241.

Sato, H., 1989. Broadening of seismogram envelopes in the randomly inhomogeneous lithosphere based on the parabolic approximation: southeastern Honshu, Japan. *Journal of Geophysical Research*, **94**, 17735–17747.

Sato, H., 1994. Multiple isotropic scattering model including P-S conversion for the seismogram envelope formation. *Geophysical Journal International*, **117**, 487–494.

Sato, H., 2006. Synthesis of vector wave envelopes in three-dimensional random elastic media characterized by a Gaussian autocorrelation function based on the Markov approximation: plane wave case. *Journal of Geophysical Research*, **111**, B06306.

Sato, H., 2007. Synthesis of vector wave envelopes in three-dimensional random elastic media characterized by a Gaussian autocorrelation function based on the Markov approximation: spherical wave case. *Journal of Geophysical Research*, **112**, B01301.

Sato, H., 2008. Synthesis of vector-wave envelopes in 3-D random media characterized by a nonisotropic Gaussian ACF based on the Markov approximation. *Journal of Geophysical Research*, **113**, B08304.

Sato, H., 2009. Green's function retrieval from the CCF of coda waves in a scattering medium. *Geophysical Journal International*, **179**, 1580–1583.

Sato, H., and Fehler, M., 1998. *Seismic Wave Propagation and Scattering in the Heterogeneous Earth*. New York: Springer.
Sato, H., and Korn, M., 2009. Synthesis of vector-wave envelopes in random elastic media on the basis of the markov approximation. In Sato, H., Fehler, M., and Dmowska, R. (eds.), *Earth Heterogeneity and Scattering Effect on Seismic Waves*. New York: Academic. Advances in Geophysics, Vol. 50, pp. 43–93.
Sato, H., and Nishino, M., 2002. Multiple isotropic-scattering model on the spherical Earth for the synthesis of Rayleigh-wave envelopes. *Journal of Geophysical Research*, **107**, 2343.
Sato, H., Fehler, M., and Dmowska, R. (eds.), 2008. *Earth Heterogeneity and Scattering Effects on Seismic Waves*. Amsterdam: Academic. Advances in Geophysics, Vol. 50.
Sawazaki, K., Sato, H., Nakahara, H., and Nishimura, T., 2009. Time-Lapse changes of seismic velocity in the shallow ground caused by strong ground motion shock of the 2000 Western-Tottori Earthquake, Japan, as revealed from coda deconvolution analysis. *Bulletin Seismological Society of America*, **99**, 352–366.
Scherbaum, F., and Sato, H., 1991. Inversion of full seismogram envelopes based on the parabolic approximation: estimation of randomness and attenuation in southeast Honshu, Japan. *Journal of Geophysical Research*, **96**, 2223–2232.
Sens-Schönfelder, C., and Wegler, U., 2006a. Radiative transfer theory for estimation of the seismic moment. *Geophysical Journal International*, **167**, 1363–1372.
Sens-Schönfelder, C., and Wegler, U. C., 2006b. Passive image interferometry and seasonal variations of seismic velocities at merapi volcano, indonesia. *Geophysical Research Letters*, **33**, L21302.
Shapiro, S. A., and Kneib, G., 1993. Seismic attenuation by scattering: theory and numerical results. *Geophysical Journal International*, **114**, 373–391.
Shearer, P., and Earle, P., 2004. The global short-period wavefield modelled with a Monte Carlo seismic phonon method. *Geophysical Journal International*, **158**, 1103–1117.
Shearer, P., and Earle, P., 2008. Observing and modeling elastic scattering in the deep Earth. In Sato, H., Fehler, M., and Dmowska, R. (eds.), *Earth Heterogeneity and Scattering Effect on Seismic Waves*. New York: Academic. Advances in Geophysics, Vol. 50, pp. 167–193.
Sheng, P., 1995. *Introduction to Wave Scattering, Localization and Mesoscopic Phenomena*. San Diego: Academic.
Shiomi, K., Sato, H., and Ohtake, M., 1997. Broad-band power-law spectra of well-log data in Japan. *Geophysical Journal International*, **130**, 57–64.
Snieder, R., 2002. Coda wave interferometry and the equilibration of energy in elastic media. *Physical Review E*, **66**, 46615.
Snieder, R., 2004a. Coda wave interferometry. In McGraw-Hill Yearbook of Science & Technology. New York: McGraw-Hill, pp. 54–56.
Snieder, R., 2004b. Extracting the Greens function from the correlation of coda waves: a derivation based on stationary phase. *Physical Review E*, **69**, 46610.
Snieder, R., 2006. The theory of coda wave interferometry. *Pure and Applied Geophysics*, **163**, 455–473.
Snieder, R., and Page, J., 2007. Multiple scattering in evolving media. *Physics Today*, **60**, 49–55.
Snieder, R., Grêt, A., Douma, H., and Scales, J., 2002. Coda wave interferometry for estimating nonlinear behavior in seismic velocity. *Science*, **295**, 2253–2255.
Takahashi, T., Sato, H., Nishimura, T., and Obara, K., 2007. Strong inhomogeneity beneath Quaternary volcanoes revealed from the peak delay analysis of S-wave seismograms of microearthquakes in northeastern Japan. *Geophysical Journal International*, **168**, 90–99.
Takahashi, T., Sato, H., Nishimura, T., and Obara, K., 2009. Tomographic inversion of the peak delay times to reveal random velocity fluctuations in the lithosphere: method and application to northeastern Japan. *Geophysical Journal International*, **178**, 1437–1455.
Vidale, J., and Earle, P., 2000. Fine-scale heterogeneity in the Earth's inner core. *Nature*, **404**, 273–275.
Wegler, U., 2004. Diffusion of seismic waves in a thick layer: theory and application to Vesuvius volcano. *Journal of Geophysical Research*, **109**, 07303.
Wegler, U., and Lühr, B., 2001. Scattering behaviour at Merapi volcano (Java) revealed from an active seismic experiment. *Geophysical Journal International*, **145**, 579–592.
Wegler, U., Korn, M., and Przybilla, J., 2006a. Modeling full seismogram envelopes using radiative transfer theory with Born scattering coefficients. *Pure and Applied Geophysics*, **163**, 503–531.
Wegler, U., Lühr, B., Snieder, R., and Ratdomopurbo, A., 2006b. Increase of shear wave velocity before the 1998 eruption of Merapi volcano (Indonesia). *Geophysical Research Letters*, **33**, L09303.
Wu, R. S., 1982. Attenuation of short period seismic waves due to scattering. *Geophysical Research Letters*, **9**, 9–12.
Wu, R. S., 1985. Multiple scattering and energy transfer of seismic waves-separation of scattering effect from intrinsic attenuation I. Theoretical modelling. *Geophysical Journal International*, **82**, 57–80.
Wu, R. S., 1989. Seismic wave scattering. In James, D. E. (ed.), *The Encyclopedia of Solid Earth Geophysics*. New York: Van Nostrand Reinhold, pp. 1166–1187.
Wu, R. S., and Aki, K., 1988. Multiple scattering and energy transfer of seismic waves-separation of scattering effect from intrinsic attenuation II. Application of the theory to Hindu Kush region. *Pure and Applied Geophysics*, **128**, 49–80.
Wu, R., and Flatté, S., 1990. Transmission fluctuations across an array and heterogeneities in the crust and upper mantle. *Pure and Applied Geophysics*, **132**, 175–196.
Wu, R., Xu, Z., and Li, X., 1994. Heterogeneity spectrum and scale-anisotropy in the upper crust revealed by the German continental deep-drilling (KTB) holes. *Geophysical Research Letters*, **21**, 911–914.
Yoshimoto, K., 2000. Monte Carlo simulation of seismogram envelopes in scattering media. *Journal of Geophysical Research*, **105**, 6153–6161.
Zeng, Y., 1993. Theory of scattered P-and S-wave energy in a random isotropic scattering medium. *Bulletin Seismological Society of America*, **83**, 1264–1276.
Zheng, Y., and Wu, R., 2005. Measurement of phase fluctuations for transmitted waves in random media. *Geophysical Research Letters*, **32**, L14314.
Zheng, Y., and Wu, R., 2008. Theory of transmission fluctuations in random media with a depth-dependent background velocity structure. In Sato, H., Fehler, M., and Dmowska, R. (eds.), *Earth Heterogeneity and Scattering Effect on Seismic Waves*. New York: Academic. Advances in Geophysics, Vol. 50, pp. 21–41.
Zheng, Y., Wu, R., and Lay, T., 2007. Inverting the power spectrum for a heterogeneous medium. *Geophysical Journal International*, **168**, 1005–1010.

Cross-references

Energy Partitioning of Seismic Waves
Propagation of Elastic Waves: Fundamentals
Seismic Diffraction
Seismic, Ambient Noise Correlation

SEISMIC ZONATION

Yanxiang Yu, Mengtan Gao, Guangyin Xu
Institute of Geophysics, China Earthquake Administration, Haidian District, Beijing, China

Definition

Seismic zonation. The process of subdividing the territory into regions with respect to the level of seismic hazard. The result of seismic zonation is usually presented as a map, which is based on seismic hazard map.

Introduction

Seismic zonation is useful for hazard reduction such as earthquake-resistant design of structures, risk analysis, land-use planning, etc. Many earthquake-prone countries developed seismic zonation maps. Seismic zonation map is usually revised or updated periodically with the progress in methodology and accumulation of new data. Seismic intensity (see *Earthquakes, Intensity*) or ground motion parameters such as peak ground acceleration (PGA), peak ground velocity (PGV), and spectral acceleration at specific natural period are mostly adopted in seismic zonation map. In the early stage, most seismic zonation maps were in terms of intensity, but since the 1980s, ground motion parameters have become popular. More commonly, such maps take the exceeding probability of 10% within 50 years (return period 475 years) as standard.

Methodology

The basic method to develop seismic zonation map is the seismic hazard analysis (see *Seismic Hazard*) approach. Both deterministic and probabilistic approaches are adopted. Although some new deterministic approaches based on the computation of synthetic seismograms are used in recent years in developing seismic zonation maps (Parvez et al., 2003), nevertheless probabilistic approaches are more popular. The seismic zonation maps, based on seismic hazard maps (see *Seismology, Global Earthquake Model*), of many countries are developed by the use of probabilistic approach. The well-known USGS National Seismic Hazard Maps (Algermissen and Perkins, 1976; Algermissen et al., 1990; Frankel et al., 1996; Frankel et al., 2002; Petersen et al., 2008) and the Global Seismic Hazard Assessment Programme (GSHAP) are the typical (see *Seismic Hazard*) ones. In this article, only probabilistic method is briefly introduced.

The probabilistic method (usually referred to as PSHA – probabilistic seismic hazard analysis) was first introduced by Cornell (1968) and since then widely adopted and modified (McGuire, 1978; Bender and Perkins, 1982). There are four basic steps for assessment of PSHA (Figure 1):

Step 1: Definition of seismic sources. Sources may range from small faults to large seismotectonic provinces with uniform seismicity. The type of seismic source can be both line or area sources.

Step 2: Definition of seismicity recurrence characteristic for the sources, where each source is described by an earthquake probability distribution, or recurrence relationship. A recurrence relationship indicates the chance of an earthquake of a given magnitude to occur anywhere inside the source during a specified period of time. An upper bound earthquake is chosen for each source, which represents the maximum event to be considered.

Step 3: Development of ground motion attenuation relationships. This is usually done empirically from strong motion records (see *Earthquakes, Strong-Ground Motion*).

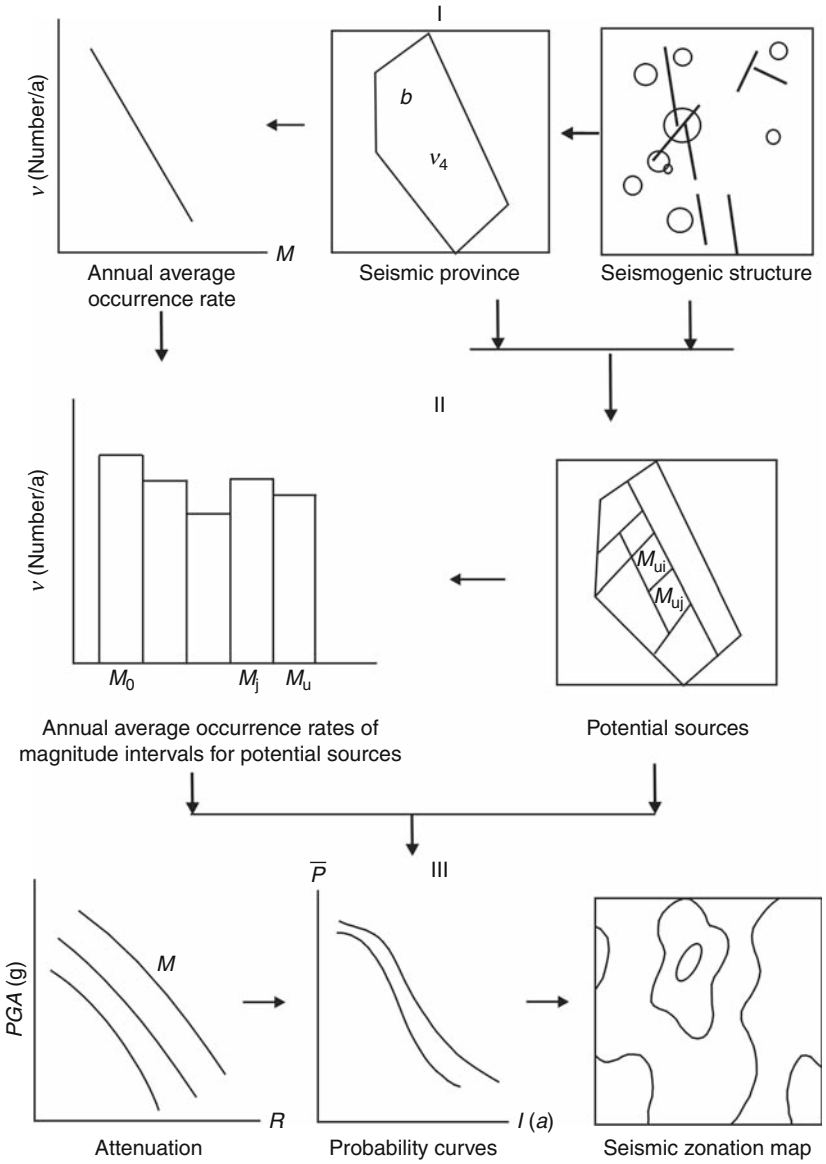
Step 4: Determination of the hazard at each site. In this case, the effects of all the earthquakes of different sizes occurring at different locations in different earthquake sources at different probabilities of occurrence are integrated into one curve that shows the probability of exceeding different levels of ground motion level (such as PGA) at the site during a specified period of time.

There are two basic assumptions in the seismic hazard analysis method at present: (a) Seismicity in the region around the site in the past indicates that in future, the recurrence rate of given site is the same as that of historic period. (b) Seismicity of the region can be expressed by tectonic earthquakes in the region, i.e., the seismic activities distribute homogeneously in a certain tectonic area or an active fault.

The two assumptions accord with the two principles of historic earthquake repeatness and tectonic extrapolation. The difference is that the PSHA adds a new concept of magnitude interval recurrent rate and the hazard is evaluated with probabilistic analysis method. The method can provide the exceedence probability of different ground motion extent (intensity, acceleration, etc.) at the site in specific time intervals, so that the earthquake resistant parameters can be selected with different exposure period, risk level, and various engineering structures.

Example: seismic zonation map of china (2001)

The first seismic zonation map of China was compiled by Wong (1921) after the Haiyuan earthquake, which occurred in 1920. After that, three versions of seismic zonation map were developed in 1957, 1977, and 1990 (Shi et al., 1992). The 1957 version demonstrated the maximum affected intensity of China. The 1977 version was provided by using the methodology of long-term and middle-term earthquake prediction. This version of seismic zonation map demonstrated the maximum encountered intensity in the forthcoming 100 years. This map was adopted by the building code. The 1990 version used probabilistic method. The seismic intensity with exceeding probability of 10% within 50 years was given in the map. It was used in the building code and other regulations related with seismic design. In 2001, a new seismic zonation map of China was issued. This map also used probabilistic method. The PGA and characteristic



Seismic Zonation, Figure 1 Approaches of probabilistic seismic hazard analysis.

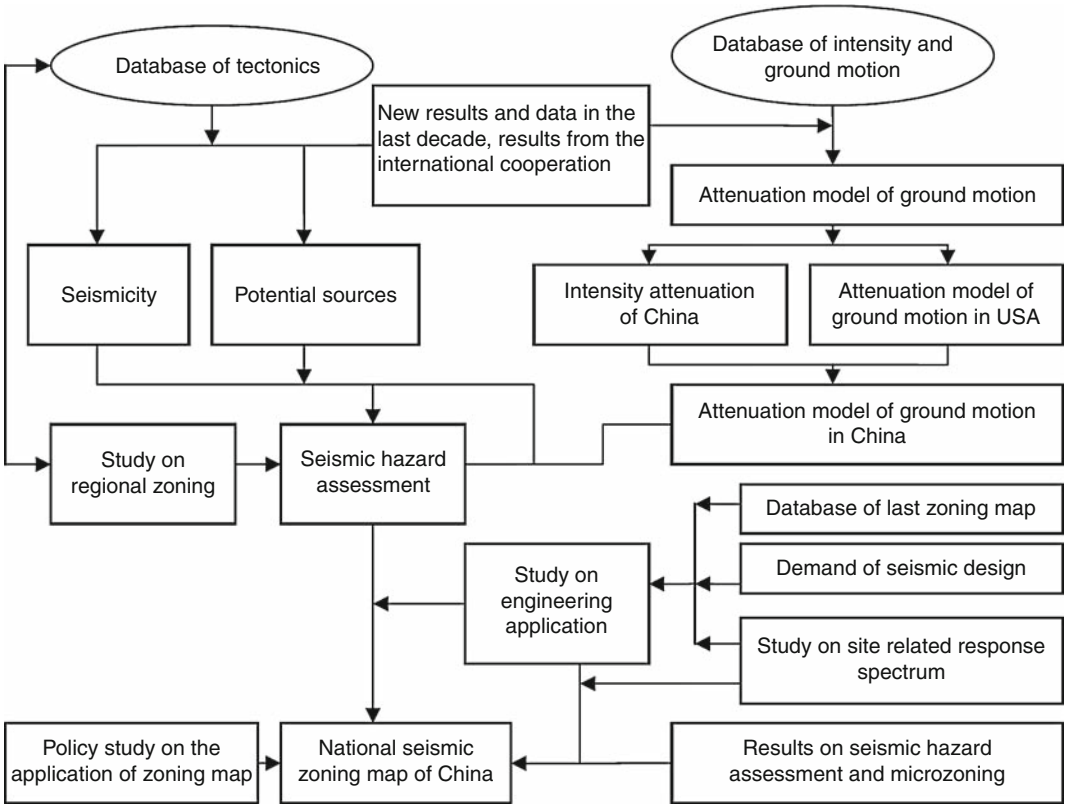
period of response spectrum Tg with exceeding probability of 10% within 50 years were given in the map. The probabilistic method used in developing seismic zonation of China (2001) was a little bit different from PSHA.

Most earthquakes in China are intraplate earthquakes (see *Seismicity, Intraplate*). Their non-homogeneity in space, non-stationarity in time were shown in the historical earthquake catalogs. The PSHA method applied in China can reveal non-homogeneity in space, non-stationarity in time. The approaches applied differ from PSHA in two aspects:

1. Evaluating seismicity and determining magnitude distribution, total annual occurrence rate of the province are on the basis of the seismic tendency estimation and the seismic characteristics analysis.
2. Annual rates of sources in all magnitude intervals are determined by spatial distribution functions, which describe relative risk among sources in the province.

The technical approach developing the national seismic zonation map of China (2001) is shown in Figure 2 (Gao, 2003).

The key scientific problems in compiling this map are: (a) how to treat the uncertainties in the evaluation of seismicity parameters and the delineation of potential sources as well as the attenuation relationship; (b) how to select the suitable parameters in the zonation map to fit the need of seismic design; and (c) how to use the domestic data and the data from the world to get the attenuation of the ground motion parameters.



Seismic Zonation, Figure 2 Technical approach of developing national seismic zonation map of China (2001).

The logic tree method is adopted to treat the uncertainties in seismic hazard analysis. The multi-set of potential source delineations was used. Four groups of scientists participated in the work of potential sources delineations based on the independent background and database. There are four sets of potential sources adopted in the logic tree analysis. Figure 3 is one of the potential source set.

Site-related seismic response spectrum is the base of seismic design for ordinary industry and civil construction. The site-related response spectrum is not only closely related to the earthquake environment, but also related to the soil condition. It is very difficult to decide site-related response spectrum by a single parameter such as intensity or peak acceleration. According to the results form seismic hazard assessment and the building code, the basic parameters used in the seismic zoning map to provide site-related response spectrum are PGA and the characteristic period of the response spectrum.

The attenuation relationships for the acceleration response spectrum platform value Sa and the velocity response spectrum platform value Sv were developed by modifying corresponding attenuation relations in the western United States according to the differences of intensity attenuation relations. Then the PGA and the characteristic period of response spectrum of acceleration Tg were defined as:

$$PGA = \frac{Sa}{2.5}$$
$$Tg = 2\pi \frac{Sv}{Sa} \quad (1)$$

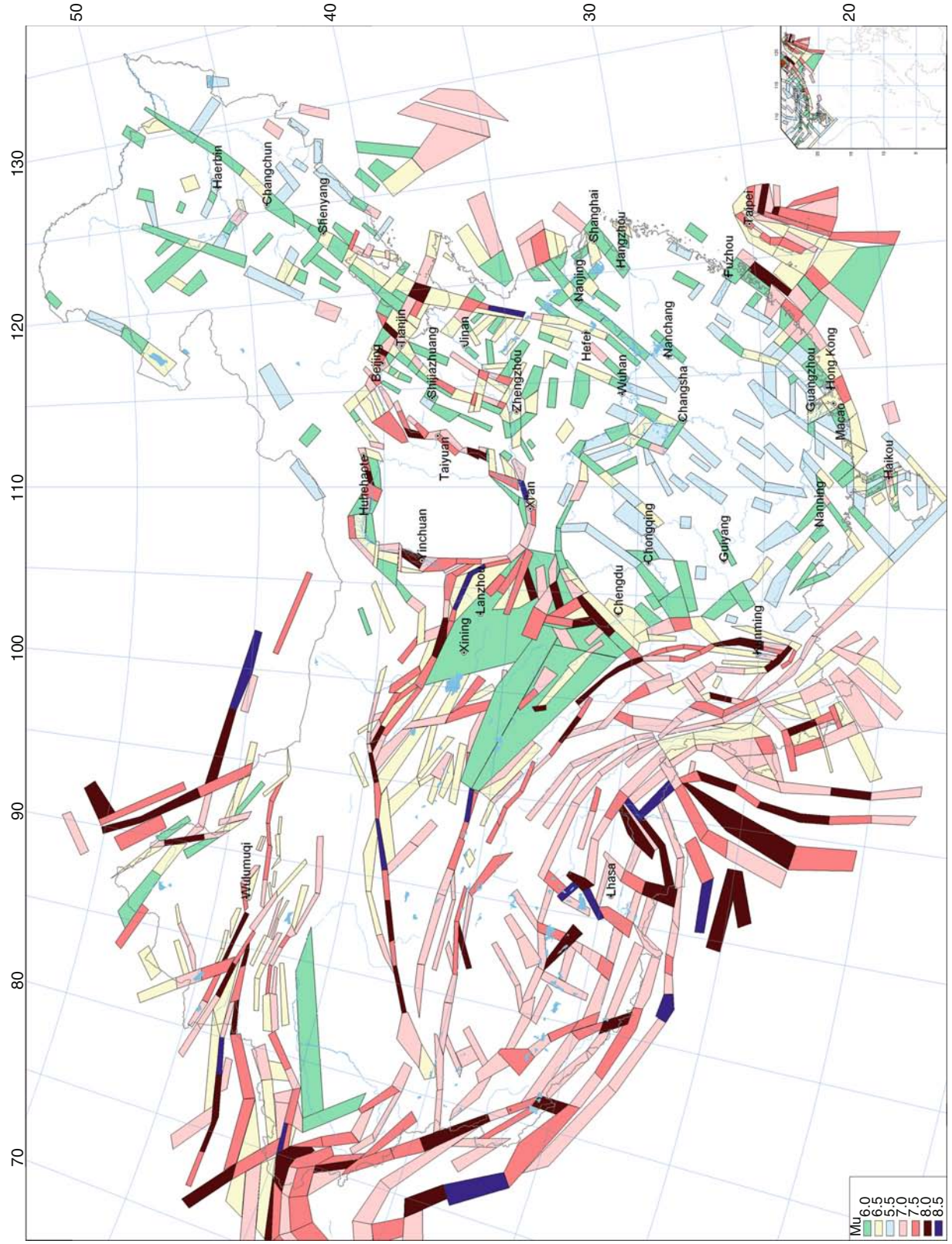
The country was divided into 40,000 grids. The probabilistic seismic hazard analysis for every grid was performed. The ground motion parameters with exceeding probability 10% within 50 years (return period 475 years) were determined.

The new zoning map includes two specific maps. One is PGA in gravity unit g (Figure 4) and the other is Tg in second (Figure 5). The scale of the maps are 1:4,000,000. The soil condition is medium hard soil. From these two parameters, the design response spectrum can be determined easily by (Equation 2):

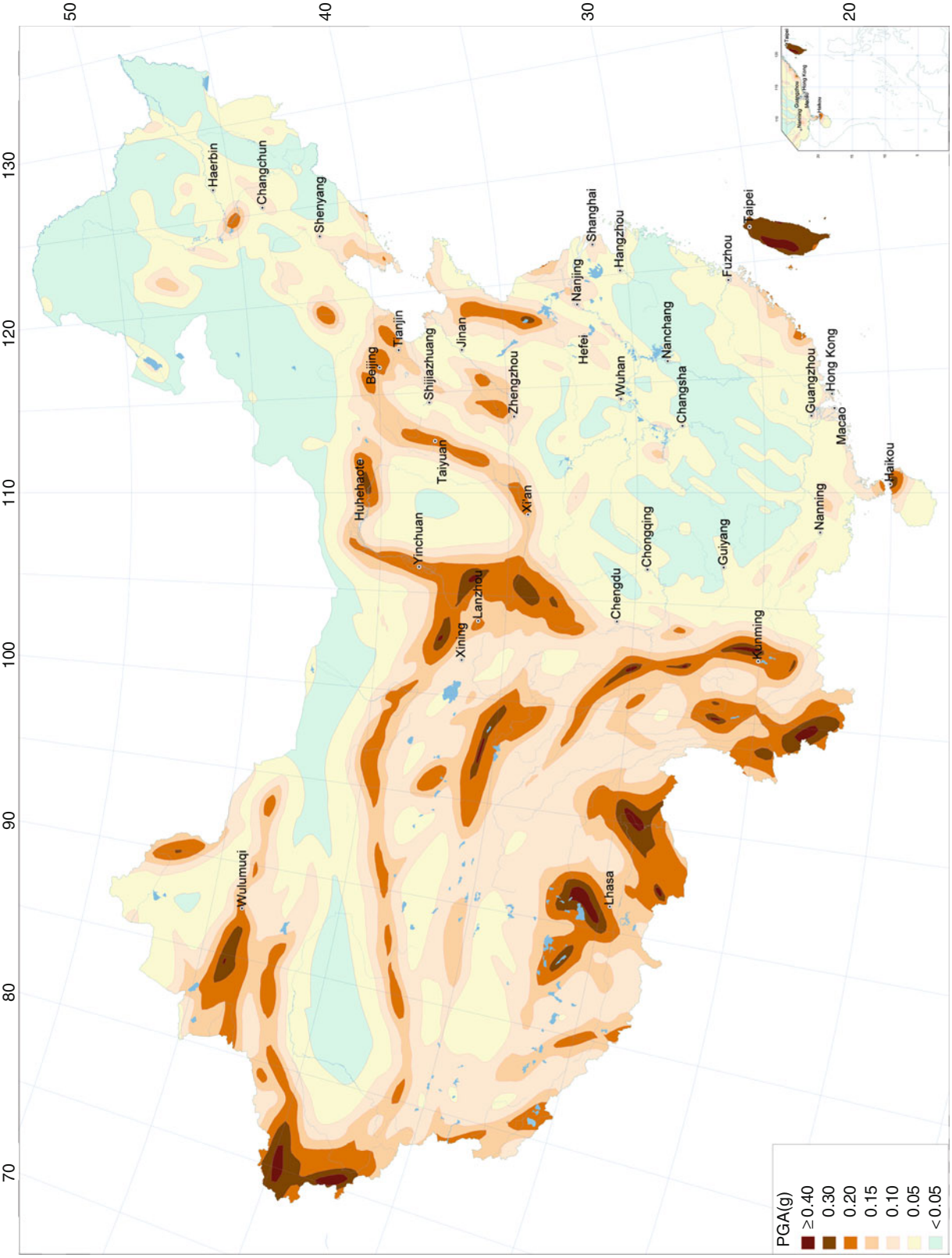
$$Sa(T) = 2.5a \begin{cases} 6T + 0.4 & 0 \le T < 0.1 \\ 1 & 0.1 \le T < Tg \\ Tg/T & T \ge Tg \end{cases} \quad (2)$$

where a is PGA in g, T is the natural period in second.

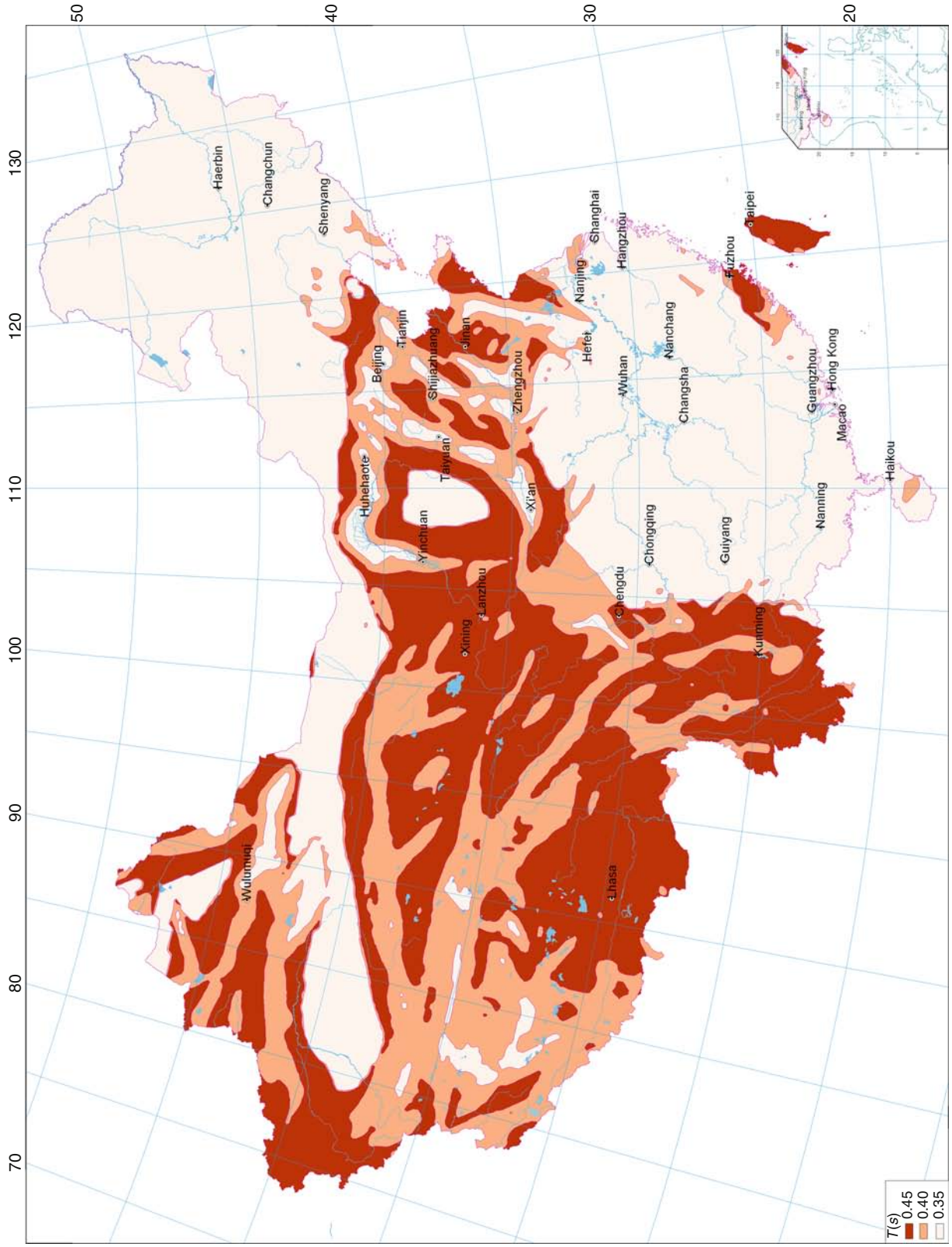
In the acceleration map (Figure 4), the territory is divided into seven zones. The acceleration for the seven zones are 0.05, 0.05, 0.10, 0.15, 0.20, 0.30, and $\geq$0.40 g, respectively.



Seismic Zonation, Figure 3 One of the seismic potential source set used in developing national seismic zonation of China (2001).



Seismic Zonation, Figure 4 Illustration of acceleration zonation map of China.



Seismic Zonation, Figure 5 Illustration of characteristic period of seismic response spectrum zonation map of China.

Seismic Zonation, Table 1 Tg value for different soil conditions (unit: second)

Zone	Soil conditions			
	Hard	Medium hard	Medium soft	Soft
1	0.25	0.35	0.45	0.65
2	0.30	0.40	0.55	0.75
3	0.35	0.45	0.65	0.90

In the characteristic period zoning map (Figure 5), the territory of China is divided into three zones. Zone 1 is the region with Tg = 0.35 s, zone 2 with Tg = 0.40 s, and zone 3 with Tg = 0.45 s.

For different soil conditions, the Tg value should be adjusted as Table 1, but the PGA value does not change with soil conditions.

The new national seismic zonation map of China serves as the obliged state standard, which took into effect in August 1, 2001. For the ordinary new constructed buildings, the standard must be followed.

Summary

Seismic zonation is a process of estimation of the seismic hazard in terms of parameters of ground motion for a certain area. Assessment results in seismic zonation map compilation, which reflects territorial distribution of the seismic hazard (see *Seismic Hazard*). Seismic zonation map is useful for hazard reduction such as earthquake-resistant design of structures, risk analysis, land-use planning, etc. Many countries apply seismic hazard maps in anti-seismic codes. Recently, the probabilistic seismic hazard analysis method is more commonly used in compiling seismic zonation map. The seismic zonataion maps take the exceeding probability of 10% within 50 years (return period 475 years) as standard. Seismic zonation will develop with the development of seismic hazard assessment methodology and anti-seismic policy.

Bibliography

Algermissen, S. T., and Perkins, D. M., 1976. A probabilistic estimate of the maximum acceleration in rock in the contiguous United States: *U.S. Geological Survey Open-File Report* 76–416, 45 pp, 2 pls., scale 1:7,500,000.

Algermissen, S. T., Perkins, D. M., Thenhaus, P. C., Hanson, S. L., and Bender, B. L., 1990. Probabilistic earthquake acceleration and velocity maps for the United States and Puerto Rico: *U.S. Geological Survey Miscellaneous Field Studies Map MF-2120*, 2 sheets, scale 1:7,500,000.

Bender, B. K., and Perkins, D. M., 1982. SEISRISK II: a computer program for seismic hazard estimation. *U.S. Geological Survey Open-File Report,* 82–293.

Cornell, C. A., 1968. Engineering seismic risk analysis. *Bulletin. Seismological Society of America*, **58**, 1583–1606.

Frankel, A. D., Mueller, C., Barnhard, T., Perkins, D., Leyendecker, E., Dickman, N., Hanson, S., and Hopper, M., 1996. National Seismic Hazard Maps – Documentation June 1996: *U.S. Geological Survey Open-File Report,* 96–532.

Frankel, A. D., Petersen, M. D., Mueller, C. S., Haller, K. M., Wheeler, R. L., Leyendecker, E. V., Wesson, R. L., Harmsen, S. C., Cramer, C. H., Perkins, D. M., and Rukstales, K. S., 2002. Documentation for the 2002 update of the National Seismic Hazard Maps: *U.S. Geological Survey Open-File Report* 2002–420.

Gao, M., 2003. New national seismic zoning map of China. *Acta Seismologica Sinica*, **16**, 639–645.

McGuire, R. K., 1978. FRISK: computer program for seismic risk analysis using faults as earthquake sources. *U.S. Geological Survey Open File-Report* 78–1007.

Parvez, I. A., Vaccari, F., and Panza, G. F., 2003. A deterministic seismic hazard map of India and adjacent areas. *Geophysical Journal International*, **155**, 489–508.

Petersen, M. D., Cramer, C. H., Reichle, M. S., Frankel, A. D., and Hanks, T. C., 2008. Discrepancy between earthquake rates implied by historic earthquakes and a consensus geologic source model for California. *Bulletin. Seismological Society of America*, **90**, 1117–1132.

Shi, Z., Yan, J., and Gao, M., 1992. Research on the principle and methodology of seismic zonation – results of the trials in North China. *Acta Seismologica Sinica*, **5**, 305–314.

Wong, W. H., 1921. Zhuizhi Collections. Beijing: The Commercial Press. (in Chinese)

Cross-references

Earthquakes, Intensity
Earthquakes, Strong-Ground Motion
Seismic Hazard
Seismic Microzonation
Seismicity, Intraplate
Seismology, Global Earthquake Model

SEISMIC, AMBIENT NOISE CORRELATION

Michel Campillo[1], Philippe Roux[1], Nikolai M. Shapiro[2]
[1]Observatoire de Grenoble, Université Joseph Fourier and CNRS, Grenoble, France
[2]Institut de Physique du Globe de Paris, Paris, France

Definition

Seismic noise: permanent motion of the Earth surface that is not related to earthquakes or specific controlled sources.

Introduction

Traditional observational methods in seismology are based on earthquake records. It results in two main shortcomings. First, most techniques are based on waves emitted by earthquakes that occurred only in geologically active areas, mainly plate boundaries. This results in a limited resolution in all other areas where earthquakes are not present. Second, the repetition of earthquakes is rare, preventing the study of continuous changes within active structures such as volcanoes or faults.

Also at smaller scales in the context of geophysics prospecting, the resolution is limited by the number and power of sources, making it difficult to image large areas and/or deep structures. Similarly, reproducible sources

are necessary for time-lapse monitoring leading to long-duration surveys that are difficult to achieve.

Nowadays, the seismic networks are producing continuous recordings of the ground motion. These huge amounts of data consist mostly of so called seismic noise, a permanent vibration of the Earth due to natural or industrial sources. Passive seismic tomography is based on the extraction of the coherent contribution to the seismic field from the cross-correlation of seismic noise between station pairs.

As described in many studies where noise has been used to obtain the Green's function between receivers, coherent waves are extracted from noise signals even if, at first sight, this coherent signal appears deeply buried in the local incoherent seismic noise. Recent studies on passive seismic processing have focused on two applications, the noise-extracted Green's functions associated to surface waves leads to subsurface imaging on scales ranging from thousands of kilometers to very short distances; on the other hand, even when the Green's function is not satisfactorily reconstructed from seismic ambient noise, it has been shown that seismic monitoring is feasible using the scattered waves of the noise-correlation function.

Theoretical basis for the interpretation of noise records at two stations

Passive seismology is an alternative way of probing the Earth's interior using noise records only. The main idea is to consider seismic noise as a wave field produced by randomly and homogeneously distributed sources when averaged over long time series. In this particular case, cross-correlation between two stations yields the Green's function between these two points. In the case of a uniform spatial distribution of noise sources, the cross-correlation of noise records converges to the complete Green's function of the medium, including all reflection, scattering, and propagation modes. However, in the case of the Earth, most of ambient seismic noise is generated by atmospheric and oceanic forcing at the surface. Therefore, the surface wave part of the Green's function is most easily extracted from the noise cross-correlations. Note that the surface waves are the largest contribution of the Earth response between two points at the surface.

Historically speaking, helioseismology was the first field where ambient-noise cross-correlation performed from recordings of the Sun's surface random motion was used to retrieve time-distance information on the solar surface. More recently, a seminal paper was published by Weaver and Lobkis (2001) that showed how, at the laboratory scale, diffuse thermal noise recorded and cross-correlated at two transducers fastened to one face of an aluminum sample provided the complete Green's function between these two points. This result was generalized to the case where randomization is not produced by the distribution of sources, but is provided by multiple scattering that takes place in heterogeneous media.

By summing the contributions of all sources to the correlation, it has been shown numerically that the correlation contains the causal and acausal Green's function of the medium. Cases of non-reciprocal (e.g., in the presence of a flow) or inelastic media have also been theoretically investigated. Derode et al. (2003) proposed to interpret the Green's function reconstruction in terms of a time-reversal analogy that makes it clear that the convergence of the noise-correlation function towards the Green's function is bonded to the stationary phase theorem. For the more general problem of elastic waves, one could summarize that the Green's function reconstruction depends on the equipartition condition of the different components of the elastic field. In other words, the emergence of the Green's function is effective after a sufficient self-averaging process that is provided by random spatial distribution of the noise sources when considering long-time series as well as scattering (e.g., Gouédard et al., 2008 and references herein).

Applications in seismology

For the first time, Shapiro and Campillo (2004) reconstructed the surface wave part of the Earth response by correlating seismic noise at stations separated by distances of hundreds to thousands of kilometers, and measured their dispersion curves at periods ranging from 5 s to about 150 s. Then, a first application of passive seismic imaging in California (e.g., Shapiro et al., 2005; Sabra et al., 2005) appeared to provide a much greater spatial accuracy than for usual active techniques. More recently, the feasibility of using the noise cross-correlations to monitor continuous changes within volcanoes and active faults was demonstrated (e.g., Brenguier, 2008a, b). These results demonstrated a great potential of using seismic noise to study the Earth interior at different scales in space and time. At the same time, the feasibility of both noise-based seismic imaging and monitoring in every particular case depends on spatio-temporal properties of the available noise wavefield. Therefore, a logical initial step for most of noise-based studies is to characterize the distribution of noise sources. Also, in many cases, knowledge of the distribution of the noise sources can bring very important information about the coupling between the Solid Earth with the Ocean and the Atmosphere. So far, we can identify three main types of existing seismological applications related to noise correlations: (1) studies of spatio-temporal distribution of seismic noise sources, (2) noise-based seismic imaging, and (3) noise-based seismic monitoring.

Noise source origin and distribution

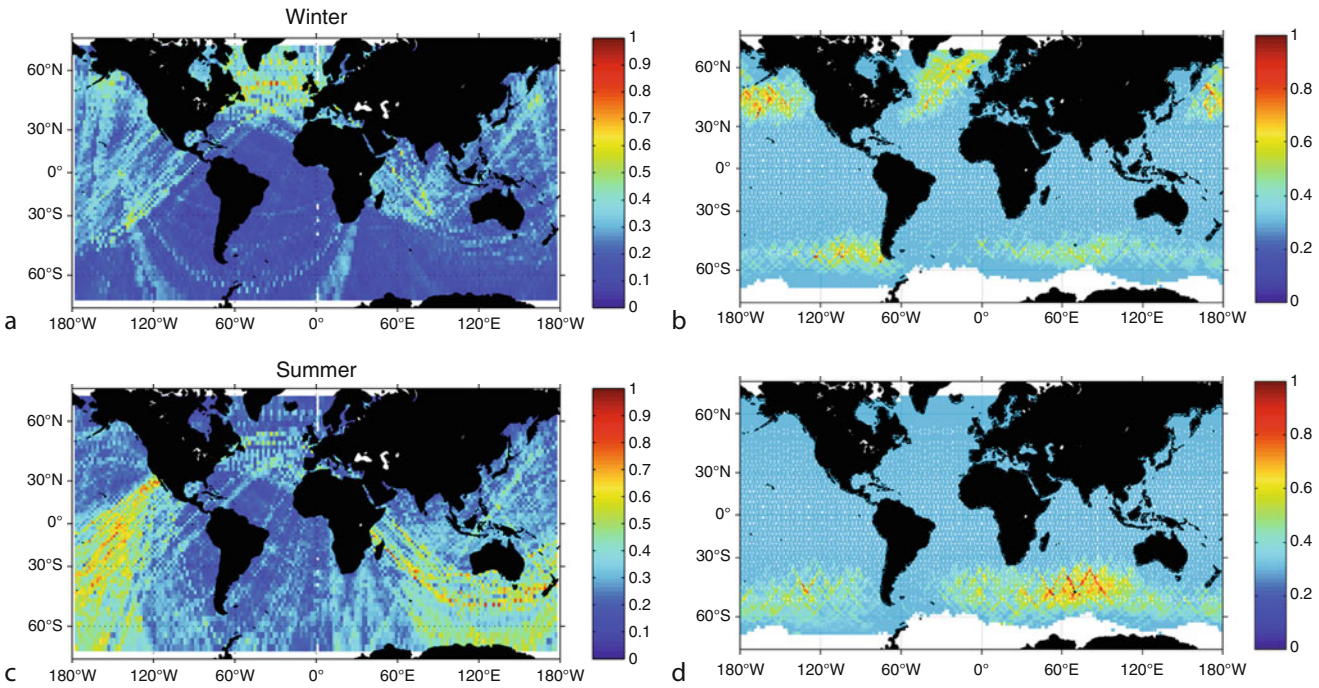
Distribution of noise sources strongly depends on the spectral range under consideration. At high frequencies (> 1 Hz), the noise is strongly dominated by local sources that may have very different origins and are often anthropogenic. At these scales, the properties of the noise wavefield should be studied separately for every particular case and no reasonable generalization can be done. At

longer periods, noise is dominated by natural sources. In particular, it is well established that two main peaks in the seismic noise spectra in so-called microseismic band (1–20 s) are related to forcing from oceanic gravity waves. It has been also argued that at periods longer than 20 s, the oceanic gravity and infragravity waves play a major role in the seismic noise excitation. The interaction between these oceanic waves and the solid Earth is governed by a complex non-linear mechanism (Longuet-Higgins, 1950) and, as a result, the noise excitation depends on many factors such as the intensity of the oceanic waves but also the intensity of their interferences as well as the seafloor topography (e.g., Kedar et al., 2008). Overall, the generation of seismic noise is expected to be strongly modulated by strong oceanic storms and, therefore, to have a clear seasonal and non-random pattern.

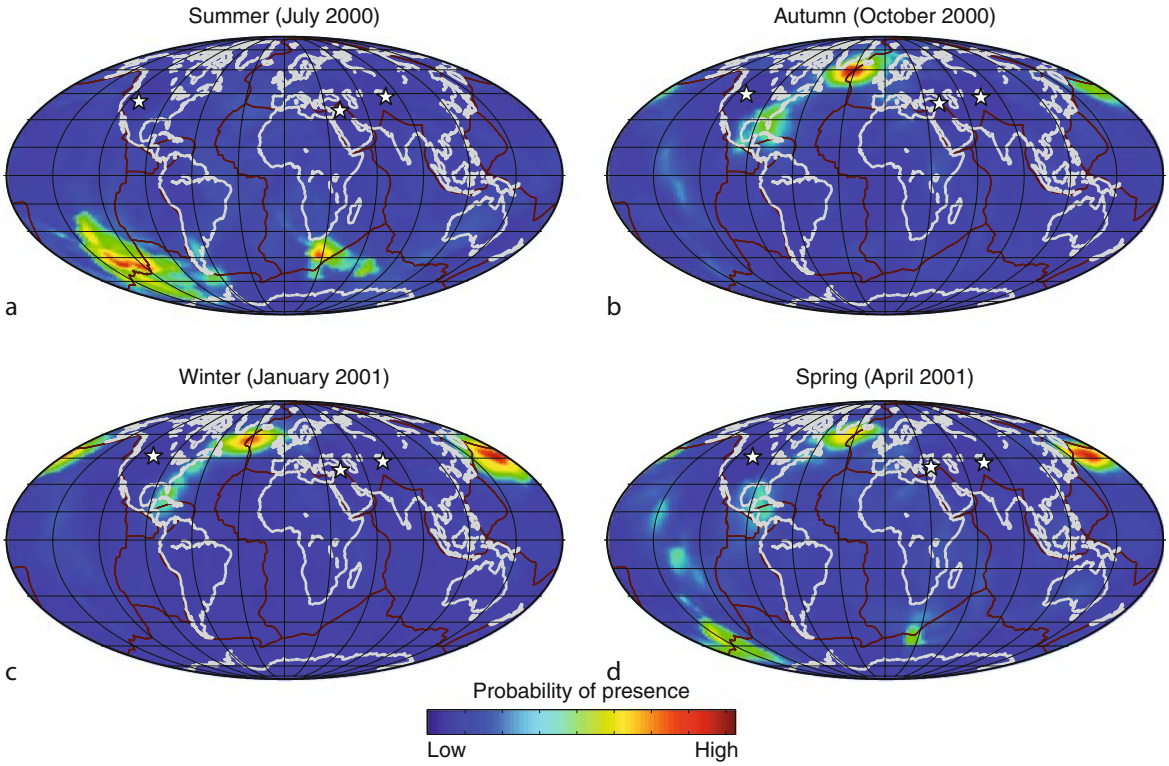
Seismic noise in the microseismic spectral band is dominated by fundamental mode surface waves. It is currently debated whether the surface wave component of microseisms is generated primarily along coastlines or if it is also generated in deep-sea areas. Inhomogeneous distribution and seasonality of microseismic noise sources is clearly revealed by the amplitude of the Rayleigh wave reconstructed in noise cross-correlations (e.g., Stehly et al., 2006) as shown in Figure 1. At the same time, body waves were detected in the secondary microseismic band and can be sometimes associated with specific storms. Figure 2 shows that sources of microseismic P waves are located in specific areas in deep ocean and exhibit strong seasonality as determined from the analysis of records by dense seismic networks (Landes et al., 2010).

Noise-based seismic imaging

Numerous studies has demonstrated that, when considered over sufficiently long times, the noise sources become sufficiently well distributed over the Earth's surface and that dispersion curves of fundamental mode surface waves can be reliably measured from correlations of seismic noise at periods between 5 and 50 s for most of interstation directions. This led to the fast development during recent years of the ambient-noise surface wave tomography. It consists of computing cross-correlations between vertical and horizontal components for all available station pairs followed by measuring group and phase velocity dispersion curves of Rayleigh and Love waves (e.g., Bensen et al., 2007). This dispersion curves are then regionalized (e.g., Lin et al., 2009) and inverted to obtain three-dimensional distribution of shear velocities in the crust and the uppermost mantle. After first results obtained in southern California (Shapiro et al., 2005; Sabra et al., 2005), this method has been applied with many regional seismological networks (e.g., Yao et al., 2006; Lin et al., 2007; Yang et al., 2008a). At smaller scales, it can be used



Seismic, Ambient Noise Correlation, Figure 1 Comparison between seasonal variations of the location of seismic noise sources and significant wave height. (**a**) and (**c**) Geographical distribution of the apparent source of the Rayleigh waves detected in the 10–20 s noise cross correlations during the winter and the summer, respectively. (**b**) and (**d**) Global distribution of the square of wave height measured by TOPEX/Poseidon during the winter and the summer, respectively (From Stehly et al., 2006).

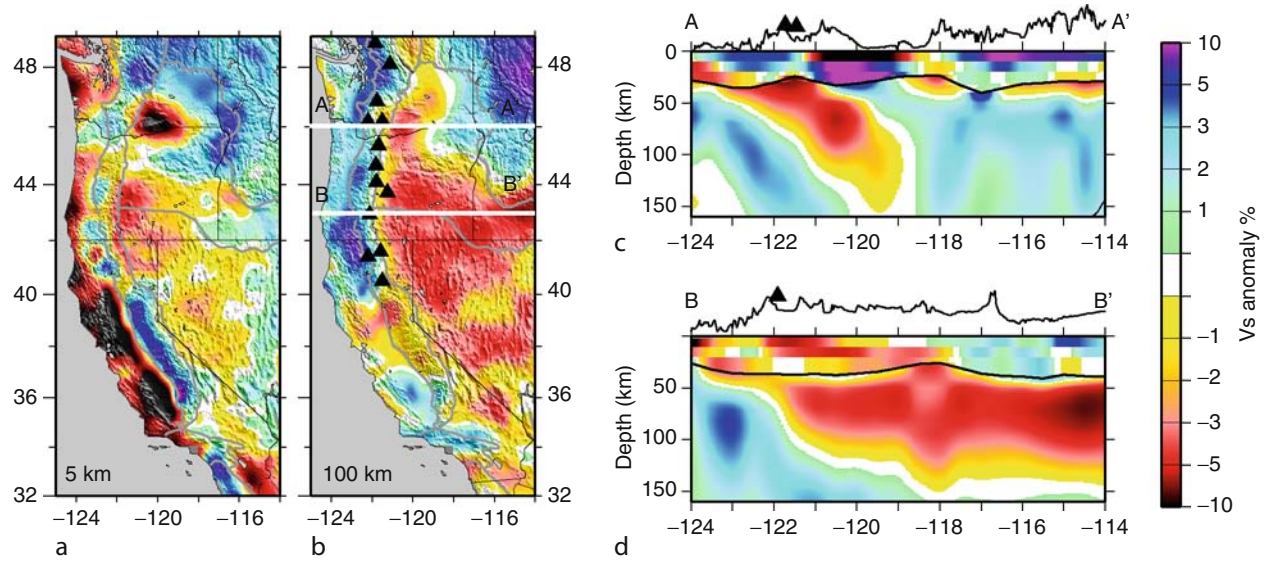


Seismic, Ambient Noise Correlation, Figure 2 Seasonal variation of the location of P-wave seismic noise sources in the secondary microseismic band (0.1–0.3 Hz) determined from the analysis of records at the three seismic networks indicated with white stars (From Landes et al., 2010).

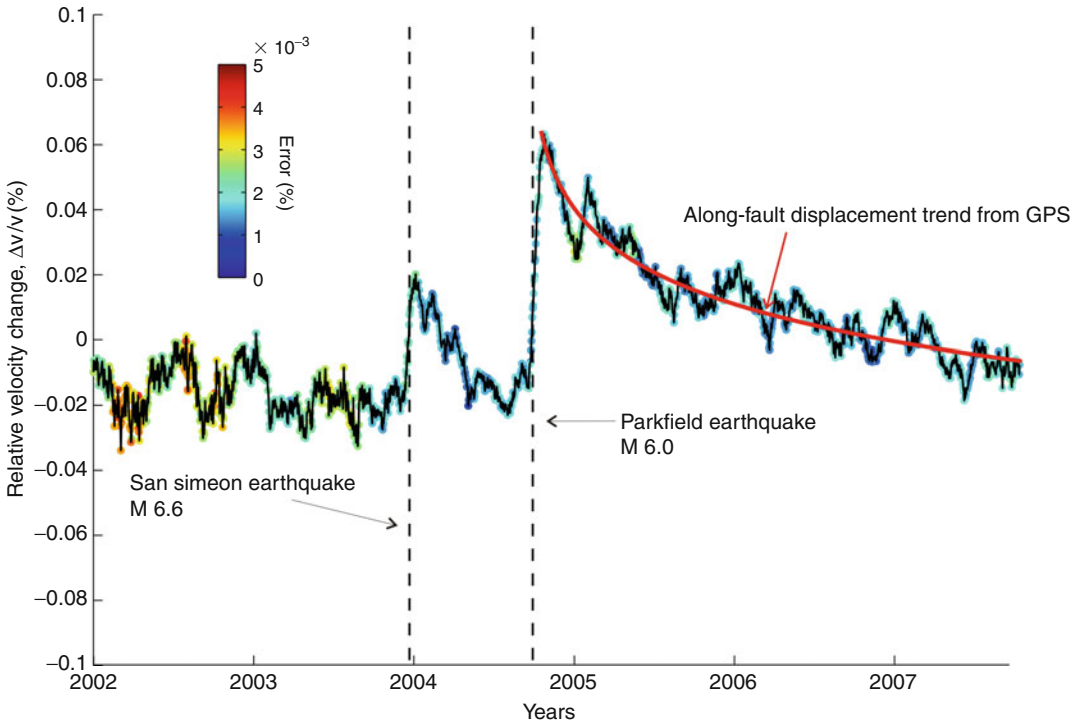
to study shallow parts of volcanic complexes (e.g., Brenguier et al., 2007). The ambient-noise surface wave tomography is especially advantageous in context of dense continent-scale broadband seismic networks such as available in USA (e.g., Moschetti et al., 2007; Yang et al., 2008b) and Europe (e.g., Stehly et al., 2009). At these scales, noise-based imaging can be used to obtain high-resolution information about the crustal and the upper mantle structure including seismic anisotropy (e.g., Moschetti et al., 2010) and can be easily combined with earthquake-based measurements to extend the resolution to larger depths (e.g., Yang et al., 2008b). An example of results obtained from combined noise and earthquakes based surface wave tomography in western USA is shown in Figure 3.

Noise-based monitoring

One of the advantages of using continuous noise records to characterize the earth materials is that a measurement can easily be repeated. This led recently to the idea of a continuous monitoring of the crust based on the measurements of wave speed variations. The principle is to apply a differential measurement to correlation functions, considered as virtual seismograms. The technique developed for repeated earthquakes (doublets), proposed by Poupinet et al., 1984, can be used with correlation functions. In a seismogram, or a correlation function, the delay accumulates linearly with the lapse time when the medium undergoes a homogeneous wave speed change, and a slight change can be detected more easily when considering late arrivals. It was therefore reasonable, and often necessary, to use coda waves for the measurements of temporal changes. Noise-based monitoring relies on the autocorrelation or cross-correlation of seismic noise records (Sens-Schönfelder and Wegler, 2006; Brenguier et al., 2008a, b). When data from a network are available, using cross-correlation take advantage of the number of pairs with respect to the number of stations. It is worth noting that the use of the coda of the correlation functions is also justified by the fact that its sensitivity to changes in the origin of the seismic noise is much smaller than the sensitivity of the direct waves. Several authors noted that an anisotropic distribution of sources leads to small errors in the arrival time of the direct waves, which can be evaluated quantitatively (e.g., Weaver et al., 2009). While in most of the cases, they are acceptable for imaging, they can be larger than the level of precision required when investigating temporal changes. The issue of the nature of the tail (coda) of the cross-correlation function is therefore fundamental and was analyzed by Stehly et al. (2008). These authors showed that it contains at least partially the coda of the Green function, i.e., physical arrivals which



Seismic, Ambient Noise Correlation, Figure 3 Shear-velocity structure of the crust and the upper mantle obtained from the inversion of the USArray data. (**a**) and (**b**) Horizontal cross-sections at depths of 5 and 100 km. (**c**) and (**d**) Vertical cross-sections along profiles delineated by the white lines in (**b**). Black lines outline the Moho. Topography is superimposed above individual cross sections. The black triangles represent active volcanoes in the Cascade Range (From Yang et al., 2008b).



Seismic, Ambient Noise Correlation, Figure 4 Relative seismic velocity change during 6 years measured from continuous noise correlations in Parkfield. The dashed lines indicated two major earthquakes: the San Simeon event that occurred 80 km from Parkfield and the local Parkfield event (Modified from Brenguier et al., 2008b).

kinematics is controlled by the wave speeds of the medium. It can therefore be used for monitoring temporal changes. As an illustration of the capability of this approach, we present in Figure 4 a measure of the average wave speed change during a period of 6 years in the region of Parkfield, California. Two main events occurred in this region during the period of study: the 2003 San Simeon and 2004 Parkfield earthquakes. In both cases, noise-based monitoring indicates a co-seismic speed drop. The measured relative variations of velocity before de San Simeon earthquake are as small as 10^{-4}. The changes of velocity associated with earthquakes are associated with at least two different physical mechanisms: (1) the damage induced by the strong ground motions in shallow layers and fault zone, as illustrated by the co-seismic effect of the distant San Simeon event, and (2) co-seismic bulk stress change followed by the post-seismic relaxation, as shown with the long-term evolution after the local Parkfield event, similar in shape to the deformation measured with GPS.

Summary

Continuous recordings of the Earth surface motion by modern seismological networks contain a wealth of information on the structure of the planet and on its temporal evolution. Recent developments shown here make it possible to image the lithosphere with noise only and to detect temporal changes related to inner deformations.

Bibliography

Bensen, G. D., Ritzwoller, M. H., Barmin, M. P., Levshin, A. L., Lin, F., Moschetti, M. P., Shapiro, N. M., and Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, **169**, 1239–1260, doi:10.1111/j.1365-246X.2007.03374.x, 2007.

Brenguier, F., Shapiro, N. M., Campillo, M., Nercessian, A., and Ferrazzini, V., 2007. 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations. *Geophysical Research Letters*, **34**, L02305, doi:10.1029/2006GL028586.

Brenguier, F., Shapiro, N., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O., and Nercessian, A., 2008a. Toward forecasting volcanic eruptions using seismic noise. *Nature Geoscience*, **1**(2), 126–130.

Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M., and Larose, E., 2008b. Postseismic relaxation along the San Andreas fault in the Parkfield area investigated with continuous seismological observations. *Science*, **321**(5895), 1478–1481.

Derode, A., Larose, E., Tanter, M., de Rosny, J., Tourin, A., Campillo, M., and Fink, M., 2003. Recovering the Green's function from field-field correlations in an open scattering medium. *The Journal of the Acoustical Society of America*, **113**, 2973–2976.

Gouédard, P., Stehly, L., Brenguier, F., Campillo, M., de Verdière Colin, Y., Larose, E., Margerin, L., Roux, P., Sanchez-Sesma, F. J., Shapiro, N. M., and Weaver, R. L., 2008. Cross-correlation of random fields: mathematical approach and applications. *Geophysical Prospecting*, **56**, 375–393.

Kedar, S., Longuet-Higgins, M., Webb, F., Graham, N., Clayton, R., and Jones, C., 2008. The origin of deep ocean microseisms in the North Atlantic Ocean. *Royal Society of London Proceedings Series A*, **464**, 777–793, doi:10.1098/rspa.2007.0277.

Landes, M., Hubans, F., Shapiro, N. M., Paul, A., and Campillo, M., 2010. Origin of deep ocean microseims by using teleseismic body waves. *Journal of Geophysical Research*, doi:10.1029/2009JB006918.

Lin, F., Ritzwoller, M. H., Townend, J., Savage, M., and Bannister, S., 2007. Ambient noise Rayleigh wave tomography of New Zealand. *Geophysical Journal International*, doi:10.1111/j.1365-246X.2007.03414.x.

Lin, F.-C., Ritzwoller, M. H., and Snieder, R., 2009. Eikonal tomography: surface wave tomography by phase-front tracking across a regional broad-band seismic array. *Geophysical Journal International*, **177**(3), 1091–1110.

Longuet-Higgins, M. S., 1950. A theory of the origin of microseisms. *Philosophical Transactions of the Royal Society of London Series A*, **243**, 1–35.

Moschetti, M. P., Ritzwoller, M. H., and Shapiro, N. M., 2007. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps. *Geochemistry, Geophysics, Geosystems*, **8**, Q08010, doi:10.1029/2007GC001655.

Moschetti, M. P., Ritzwoller, M. H., and Lin, F. C., 2010. Seismic evidence for widespread crustal deformation caused by extension in the western USA. *Nature*, **464**, 885–889, doi:10.1038/nature08951.

Poupinet, G., Ellsworth, W. L., and Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California. *Journal of Geophysical Research*, **89**, 5719–5731.

Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W. A., and Fehler, M. C., 2005. Extracting time domain Green's function estimates from ambient seismic noise. *Geophysical Research Letters*, **32**, L03310.

Sens-Schönfelder, C., and Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia. *Geophysical Research Letters*, **33**, L21302.

Shapiro, N. M., and Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, **31**, L07614, doi:10.1029/2004GL019491.

Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M., 2005. High resolution surface wave tomography from ambient seismic noise. *Science*, **307**, 1615–1618.

Stehly, L., Campillo, M., and Shapiro, N., 2006. A Study of the seismic noise from its long range correlation properties. *Journal of Geophysical research*, **111**, B10306.

Stehly, L., Campillo, M., Froment, B., and Weaver, R. L., 2008. Reconstructing Green's function by correlation of the coda of the correlation (C3) of ambient seismic noise. *Journal of Geophysical Research*, **113**, B11306.

Stehly, L., Fry, B., Campillo, M., Shapiro, N. M., Guilbert, J., Boschi, L., and Giardini, D., 2009. Tomography of the Alpine region from observations of seismic ambient noise. *Geophysical Journal International*, **178**, 338–350.

Weaver, R. L., and Lobkis, O. I., 2001. Ultrasonics without a source: thermal fluctuation correlations at MHz frequencies. *Physical Review Letters*, **87**(13), 134301, doi:10.1103/PhysRevLett.87.134301.

Weaver, R. L., Froment, B., and Campillo, M., 2009. On the correlation of non-isotropically distributed ballistic scalar diffuse waves. *Journal of the Acoustical Society of America*, 1817–1826.

Yang, Y., Li, A., and Ritzwoller, M. H., 2008a. Crustal and uppermost mantle structure in southern Africa revealed from ambient

noise and teleseismic tomography. *Geophysical Journal International*, doi:10.1111/j.1365-246X.2008.03779.x.
Yang, Y., Ritzwoller, M. H., Lin, F.-C., Moschetti, M. P., and Shapiro, N. M., 2008b. The structure of the crust and uppermost mantle beneath the western US revealed by ambient noise and earthquake tomography. *Journal of Geophysical Research*, **113**, B12310, doi:10.1029/2008JB005833.
Yao, H., van der Hilst, R. D., and de Hoop, M. V., 2006. Surface-wave array tomography in SE Tibet from ambiant seismic noise and two-station analysis – I. Phase velocity maps. *Geophysical Journal International*, **166**, 732–744.

Cross-references

Body Waves
Earthquakes and Crustal Deformation
Seismic Noise
Seismic Tomography
Seismic Waves, Scattering
Surface Waves

SEISMIC, MIGRATION

Samuel H. Gray
CGGVeritas, Calgary, AB, Canada

Synonyms

Seismic imaging

Definition

Seismic migration. A set of techniques for transforming recorded (elastic-wave) seismic reflection data into an image of reflecting boundaries in the earth's interior. In simplest form, these are intended to correct certain distortions present in recorded wavefields. The distortions can be caused by: diffractors inside the earth, which scatter the incident energy to a range of receiver locations; geometric effects caused by dipping reflectors; and velocity effects, which cause the seismic waves to change direction as they propagate from source to reflector to receiver. Migration is also used to estimate seismic velocity and to provide amplitudes for rock property analysis.

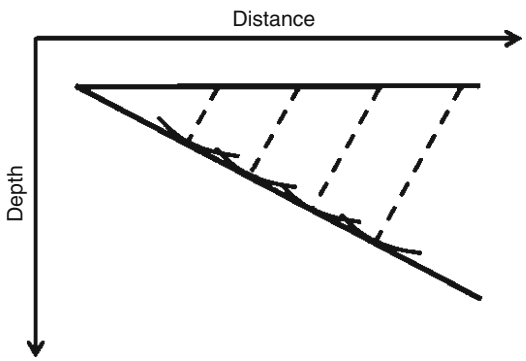
The migration heuristic assumes that subsurface reflectors are made up of point diffractors (Trorey, 1970). The response of a reflector to input seismic energy is the sum of responses to all the diffractors, and migration transforms this response to the actual reflector shape by collapsing each diffractor response to a point (diffraction stack).

History

Mapping using pencil and paper; mechanical migration

The first migrations were performed in the 1920s as a manual operation. A single-fold reflection profile was acquired with sources and receivers on the surface of the earth, and two-way travel times for events from a shallow dipping reflector were picked on the recorded traces. The extent of the lateral offset between source and receiver locations was neglected, which is equivalent to assuming that sources and receivers were colocated, or at zero offset. Then circular arc segments were drawn, centered at the source-receiver midpoints and with radii equal to one-half the product of the picked travel times and the velocity of propagation in the overburden. The envelope of these mapped arcs gave a rough picture of the reflecting interface (Figure 1).

This technique reveals principles and problems of migration. First, reflection events move sideways, or *migrate*, from their picked time locations on the recording surface (the source-receiver midpoints) to subsurface locations at the reflection interface. When reflectors in the earth are not flat, the lateral position of a subsurface reflector is different from that of the recording location of the same event. Second, the assumption of colocated sources and receivers is problematic. As the use of seismic data increased, reflection profiles were typically acquired using higher fold, or multiplicity: for each source, a spread of many receivers recorded reflection data in order to provide a greater multiplicity of observations. Some of the receivers were a considerable distance from the source locations, violating the zero-offset assumption. Later, the common-midpoint (CMP) stack (Mayne, 1962) performed approximate corrections intended to shift each event on a nonzero-offset recorded trace to its zero-offset time, once again allowing the use of the zero-offset assumption in migration (poststack migration). In areas of complicated geologic structure, however, the time corrections used in the CMP stacking process are not accurate, requiring that each recorded trace be migrated using its actual source and receiver location (prestack migration). Third, if more than one reflector exists inside the earth, the various migrated events can interfere with one another as the circular arcs that build up their envelopes overlap. This problem is solved by the more recent application of wave theory to migration. This allows the



Seismic, Migration, Figure 1 Schematic illustration of early pencil-and-paper migration. Assuming constant velocity and zero-offset source and receiver, an event on a seismic trace is swung out along an arc of a circle. Repeating this for the corresponding event on all traces produces an image of a reflector, which lies along the envelope of all the arcs.

rigorous migration of waveforms, with energy buildup near reflector locations, and energy cancellation away from reflector locations. Fourth, if the velocity used in the migration process is incorrect, the image of a subsurface reflector will be located incorrectly. This fundamental velocity problem persists today.

The process of swinging reflection events through arcs to construct approximate reflector segments spawned a number of mechanical devices that migrated reflection records. They allowed the migration of picked events from nonzero-offset traces. However, they were all based on the same principle of swinging events over a restricted range of angles. This process was formalized by Hagedoorn (1954).

Digital migration

Seismic recording, processing, and migration became digital in the 1960s. Digital processing emphasized the discrete sampling of signals, in both time and space. It also allowed reflection records to be treated as wavefields, and wavefield algorithms could be applied to them. Wave-equation migration methods arose, based on differential equations, integral solutions to the wave equation, or mapping in the Fourier domain (Berkhout, 1982; Yilmaz, 1987). These methods are in use today. They have different realms of validity and different efficiencies. One of the methods, integral (Kirchhoff) migration, deserves mention. This technique is the direct descendant of pencil-and-paper migration. In Kirchhoff migration, samples on the recorded input traces are swung out over a range of angles. The earlier techniques could consider only isolated events, because constructing the envelopes corresponding to all reflection events of a record would produce an uninterpretable mess. However, including the wave equation as an essential part of the process allowed all samples to be swung out. Constructive and destructive interference, a natural consequence of treating the samples as components of a wavefield, caused reflection events to appear on the image, not as envelopes, but as waveforms moved to reflector locations.

Poststack migration

The first digital migrations were performed on computers with limited memory and processing capability. Migrating all traces of a two-dimensional (2-D) reflection survey involving hundreds of source locations, each with tens of receiver locations, would have taken months on early computers. So it became necessary to make use of the intermediate process of CMP stack to reduce the number of traces to be migrated. As mentioned above, this process performs time ("normal moveout," or NMO) corrections to traces with nonzero source-receiver offset. These corrections are time shifts, different for each sample, intended to produce equivalent zero-offset traces to be summed, or stacked, together. The quality of the normal-moveout process depends on the stacking velocity, which determines the time shift to be applied to each sample of each nonzero-offset trace. In order to stack together traces in a domain where the velocity varies as little as possible, it is desirable to sort the input traces from their original recorded common-shot configuration to a set of records where each record has a common source-receiver midpoint – thus the term common-midpoint stack. The stacking velocity for a common-midpoint gather is taken to be the velocity at the midpoint location. This velocity can vary with time down the zero-offset trace of the record. For each value of time, it is a type of average velocity. When all the traces in a CMP gather have been NMO-corrected and summed (stacked) together, the CMP gather has been replaced with a single, zero-offset, trace. The NMO/stack process is illustrated on a single reflection event in Figure 2. The ensemble of zero-offset traces, from all CMP gathers indexed by their midpoint locations, is a stacked record ready to be (poststack) migrated.

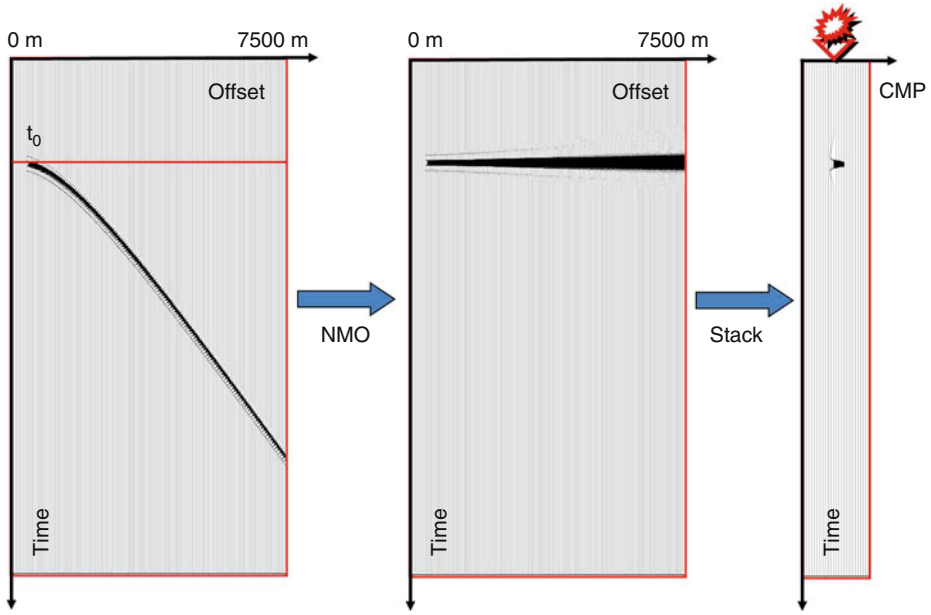
The CMP stacking process makes several approximations that are incompatible with the physics of wave propagation inside the earth. For example, sorting into CMP records and assuming that the seismic velocity does not change for all traces within a record ignores the lateral variability of the actual velocity, and it assumes that the reflectors inside the earth are locally flat below each midpoint location. These approximations were not usually damaging in earlier times, when drilling prospects were in shallow, relatively simple geology. As shallow targets were drilled, leaving deeper prospects under complex overburden with complex velocity behavior, the approximations broke down. Since about 2000, migration performed without CMP stacking (prestack migration) has become commonplace.

Prestack migration

Dropping the NMO/stack process from the migration flow eliminated some problems, but it made the velocity estimation problem more complicated. NMO velocities were useful for migration as well as stacking. These velocities were not necessarily accurate, but they provided a guide for the velocities to be used in migration. Using a rough estimate of velocity, poststack migration generates plausible migrated images, even if they lack the resolution required for interpretation of subtle targets. Accurate prestack migration, on the other hand, requires the velocity to be estimated without a prior stacking velocity analysis. New tools, consistent with wave propagation inside a geologically complicated earth, were needed for this task. These tools have been under development since the 1980s, but are not yet completely satisfactory. Unlike NMO/stack, they allow arbitrary velocity structure inside the earth, and there is no "locally flat" assumption.

2-D, 3-D, wide-azimuth 3-D acquisition and migration

Early seismic reflection profiles were acquired along 2-D lines, both land and marine. On land, a line of receivers was laid out, and shots were fired into the earth at locations along the line. As the shot locations moved along the line,



Seismic, Migration, Figure 2 NMO and CMP stack illustrated on a single reflection event.

receivers were picked up from the back end of the spread and placed at the front. Marine data were acquired using a vessel towing a linear array of receivers (a streamer), firing at regular locations along the line. 2-D processing assumes that the subsurface of the earth is invariant in the direction perpendicular to the line direction. When this assumption is violated, "out-of-plane" reflectors (not vertically below the acquisition line) contribute energy to the reflection records; 2-D migration cannot map this energy to its correct locations. For this reason, 3-D acquisition became commonplace in the 1970s and 1980s. On land, the receivers were laid out in a 3-D pattern (patch) that could be repeated as in 2-D. Gradually, the receiver patch became larger (thousands of meters on a side), with a wider range of azimuths, or compass directions of vectors joining the source and receiver locations. Today, the recording patch often contains several thousand receivers. Wide-azimuth marine acquisition is more difficult than land because it is physically impossible for a single vessel to tow the large number of streamers required. However, adding a second vessel as a source vessel alongside the first vessel at a large lateral distance produced data over a wide range of azimuths as for land.

Moving from 2-D to 3-D enabled the accurate migration of geologically complex structures such as salt bodies and overthrust structures, and it altered the economics of migration. In the initial days of 3-D seismic data, 2-D prestack migration was available but not commonly applied because it was still too computationally expensive. A fortiori, 3-D prestack migration was not yet feasible. In 3-D, the numbers of input and output locations are increased by an order of magnitude over 2-D, increasing migration cost by two orders of magnitude.

Deep crustal imaging

Most migration projects are performed by the oil and gas industry for hydrocarbon exploration, whose target depths rarely exceed 10 km. However, solid-earth geophysicists acquire and migrate seismic data to map crustal structures tens of kilometers deep. Imaging very deep targets with cumulative overburden effects such as wavefield scattering and absorption presents added challenges, and migration cannot be expected to be successful as often as it is for hydrocarbon exploration (Warner, 1987).

Time migration, depth migration

Seismic traces are recorded in time, with a particular event occurring at different times on traces with different lateral positions. The CMP stack produces a composite section, with a trace at each position. Although migration is used to image reflectors inside the earth, the CMP stack itself provides a crude image. With its collapse of many seismic records into one, and its use of redundancy (fold) to suppress noise, the CMP stack produces images that can sometimes be interpreted without migration. If the CMP stack is migrated with the depth axis on the migrated image replaced by time, events on the unmigrated stack and the migrated stack can be compared directly, helping with interpretation. Such a migration, called "time migration," is opposed to the more natural "depth migration."

This fact has caused a diversity of approaches to migration. Geophysical interpreters often prefer time migration, but structural geologists and engineers usually prefer depth migration. Structural geology deals with space (depth) and how space is filled with material; geologists and engineers are concerned with the physical dimensions

of a potential hydrocarbon trap. Even if a migrated image is interpreted in time, in principle it is preferable to perform the migration and display the final image in depth. In practice, however, there are valid arguments against this statement.

First, time migration is sometimes more robust than depth migration; that is, it is often easier to produce an interpretable image with time migration than with depth migration, even if the image is known not to be accurate. This is because time and depth migration treat velocity differently. Time migration, following procedures developed for NMO/stack, uses what is best called an imaging velocity field, i.e., one that best focuses the migrated image. At each image location, this velocity is an average of the seismic velocity values in the overburden above the image depth. The imaging velocity is free to vary spatially. In essence, time migration performs a constant-velocity migration at each image location, where the constant can change from point to point. The imaging velocity field that produces the best image need not be consistent with any possible field of actual velocities inside the earth, and any attempt to convert the imaging velocities to a field of geologically possible earth velocities can fail. For example, assuming that the imaging velocities are root-mean-square velocities and using the Dix (1955) equation to invert these to interval velocities often produces physically impossible values. This inconsistent handling of velocity allows time migration to be robust. Second, the precision implied by depth migration is not always justified. A migrated image purports to show locations of reflectors and diffractors inside the earth, but often the locations are incorrect. It is easy to see that this can happen for time migration, which is not precise because of its inconsistent handling of velocity. For depth migration, the implied accurate positioning of reflectors assumes a correct migration velocity, which is an interval velocity field, i.e., a field of local velocities. Using accurate interval velocities allows depth migration to image more accurately than time migration can, but using inaccurate interval velocities can degrade image coherence that the less precise velocities of time migration often preserve.

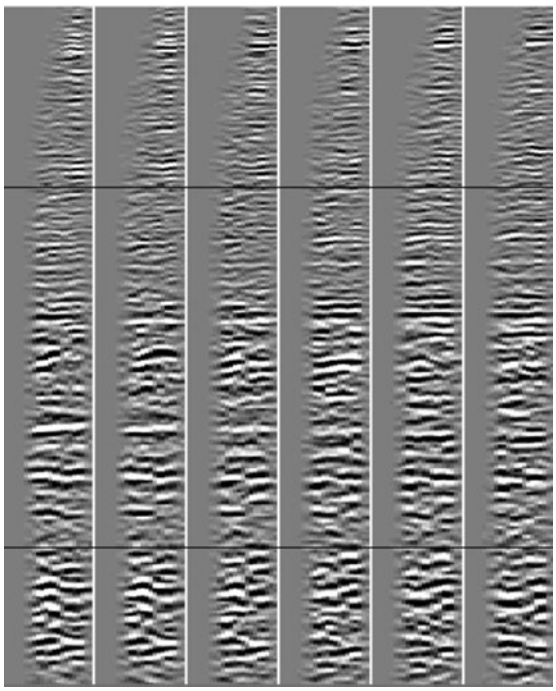
Purposes of migration

Strucural imaging

The primary goal of seismic migration is to correct distortions present in recorded seismic records, due to uncollapsed diffraction energy, geometric effects, or velocity effects. Even in areas of fairly flat geology, diffracted energy from abrupt fault terminations can obscure deeper reflection events, making their interpretation difficult. Therefore, collapsing unwanted diffraction energy to scatterer locations is important. Geometric and velocity effects are caused by complicated geologic structures. Familiar examples are "bowtie" signatures on stacked records converted by migration into synclines, and broad anticlinal structures converted by migration into narrower, steeper structures. Other examples include complicated water bottoms, complicated salt geometry, thrusted structures, heavily faulted areas, and gas-charged sediments. In these, reflector geometries or lateral velocity variations cause wavefield distortions that must be corrected before the image can be interpreted.

Velocity estimation

Poststack migration is performed using a velocity field that is specified beforehand. The velocity can be derived from NMO velocity analysis or from an assumed geologic model. It is difficult to analyze a poststack migrated image for velocity errors. By contrast, we can analyze prestack migration for velocity errors. To do this, one performs prestack migration separately on all the input records, with individual images that overlap considerably. In the overlap areas, the different images can be checked for consistency by sorting the migrated traces into common-image-point gathers (CIGs), which are the migration analog of NMO-corrected unmigrated CMP gathers. Within a CIG, each trace refers to a single surface location, imaged from a different input record. As an example, Figure 3 shows six CIG's from a land survey, referring to images below six different locations on the earth's surface. Each CIG displays migrated traces from different source-receiver offsets. If all the individual migrations have imaged one particular event at the same depth, the images are



Seismic, Migration, Figure 3 Prestack depth migration common-image gathers from six different locations on a survey from Western Canada. Each of the gathers contains traces from migrations of data with offsets ranging from 0 m to 4,500 m. The events on the gathers are flat except at the greatest depths, indicating that the velocity field is largely correct.

consistent at that location. That is the case in this example: except at the greatest depths, most of the events on each CIG appear flat. If the different migrations image a particular event at different depths, the images will be inconsistent, and the event will not appear flat on a CIG. In order to make the images consistent for such an event, the migration velocity will need to be adjusted in the overburden above the event location.

For prestack depth migration, the velocity estimation problem is to build an interval velocity field that makes the migrated events appear flat on all the CIGs. This is usually posed as an algebraic inverse problem, called seismic tomography, involving the interval velocity field, the reflector locations, and the raypaths from the reflectors to the source and receiver locations. This problem is solved iteratively for velocity model refinements. In order to avoid converging to an unrealistic model, tomography usually needs to be constrained, using partial knowledge of the seismic velocities and reflecting horizon locations. This biases the solution, and it can result in incomplete event flattening on the CIGs. Tomography is widely used, but it usually produces velocity models with limited spatial resolution

Increasingly, seismic velocities are recognized as anisotropic: at each location, seismic velocity varies with direction. Including anisotropy makes velocity estimation more difficult, but anisotropy is usually needed to produce accurate images. The most common assumption for anisotropy is transverse isotropy, often with a tilted axis of symmetry (TTI).

Currently, almost all migrations are performed before stack, and migration is used to estimate velocity (imaging velocity for time migration, interval velocity for depth migration).

Migrated amplitude analysis for rock property determination

In the 1980s, researchers observed that careful handling of amplitude terms within migration algorithms produced migrated amplitudes that are proportional to reflection coefficient values at rock boundaries. This observation is the basis of true-amplitude migration (Bleistein, 1987). Migrated amplitudes are now often analyzed in terms of angle-dependent reflection coefficients at rock boundaries. The reflection coefficients can be used to estimate the types of rock above and below each interface, and whether the rocks can bear hydrocarbons. This endeavor has had some success in areas of moderate velocity and structural complexity, but less success below areas of high complexity, such as below salt.

Imaging conditions

Isochron imaging

Pencil-and-paper migrations were performed by swinging reflection events to their migrated locations, which are the envelopes of all possible swing locations. Kirchhoff migration can be performed the same way, with the wave equation canceling the energy away from the envelopes and leaving only images of reflectors. The time of a sample on an input trace is the sum of time from the source to a reflector plus time from the reflector to the receiver; Kirchhoff migration acts by placing the sample at all subsurface locations (isochrons) that share the same two-way travel time.

Isochron imaging, useful in describing Kirchhoff migration, is plausible and correct, but it is not a wavefield concept, and cannot be applied to all migrations based on the wave equation. These require more general treatment, which is different for poststack and prestack migration.

Poststack migration – exploding reflector model

The CMP stack simulates a zero-offset record, where energy has been excited and recorded at the same location. It can also be considered as a single wavefield from a far-fetched experiment. Suppose all the reflectors and diffractors inside the earth exploded at the same instant (time zero), sending energy in all directions; some of the energy would emerge at the earth's surface at later times to be recorded at the receiver locations. Suppose also that the propagation velocity of the wavefield from the exploding reflectors is exactly one-half the actual propagation velocity inside the earth. Then the recording time for an event from an exploding reflector is the same as the two-way zero-offset reflection time, with the actual velocity, from the same reflector. This thought experiment (Loewenthal et al., 1976) provides the basis for a poststack migration imaging principle, breaking migration into two parts: downward continuation and imaging. The first part expresses the wavefield at one depth in terms of the wavefield at a shallower depth. As the wavefield moves down, time decreases towards zero (the time of the explosion). The second part picks off the downward-continued wavefield at each location when the time reaches zero. The result of the process is a snapshot of the reflectors in the act of explosion.

Reflected wavefield amplitude normalized by source wavefield amplitude

For prestack migration, downward continuation is useful, but exploding reflectors and zero-time imaging are not. As in poststack migration, the wavefield from the many receiver locations is downward continued into the earth and backwards in time. Likewise, the wavefield from the source location is downward continued, but forward in time so that it can interact with the receiver wavefield at actual reflection times. If the two downward-continued wavefields intersect at a location at a particular time, the process has captured the instant of reflection at that point. At that location and time, the receiver wavefield amplitude equals the source wavefield amplitude times the value of the reflection coefficient at the location (possibly zero, if the location is not on an actual reflector). That is, the migrated image at that location is the receiver wavefield divided by the source wavefield.

Deconvolution and cross correlation imaging conditions

Some migration techniques are applied in the time domain, others in the frequency (ω) domain. For frequency-domain methods, a precise statement of the prestack migration imaging principle is that reflectors exist inside the earth where downward continued source and receiver wavefields are in phase, and the reflection strength is the ratio of the two wavefields. For a single frequency, there is complete uncertainty about the location of a reflector; only by adding the results from many frequencies can reflector positions be precisely determined. Thus, the prestack-migration imaging principle can be written as

$$R(\mathbf{x}) = \int \frac{U(\mathbf{x};\omega)}{D(\mathbf{x};\omega)} \mathrm{d}\omega, \tag{1}$$

where $R(\mathbf{x})$ is reflection strength at location $\mathbf{x}$, U is the downward-continued receiver wavefield, and D is the downward-continued source wavefield. The quotient can cause stability problems when the source wavefield D becomes weak. To overcome instability, note that if both numerator and denominator are multiplied by the complex conjugate of D, denoted D^*, the phase of the denominator will disappear (the denominator becomes a positive real number). Instability, caused by the nearly vanishing denominator DD^*, can be eliminated by replacing the denominator with a constant value, leaving the modified imaging condition

$$R(\mathbf{x}) = \int U(\mathbf{x};\omega) D*(\mathbf{x};\omega) \mathrm{d}\omega. \tag{2}$$

Equations 1 and 2 migrate events to the same locations, but migrated amplitudes obtained using Equation 2 are different from those obtained using Equation 1. Imaging conditions (1) (*deconvolution*) and (2) (*crosscorrelation*) were introduced by Claerbout (1971).

Migration techniques

Integral (Kirchhoff) migration

Kirchhoff migration is the digital embodiment of the earliest migration heuristics, and it has a basis in the downward continuation of wavefields: the wavefield at depth is an integral of the recorded wavefield (Schneider, 1978). The mathematical physics of this operation requires amplitude and phase manipulations of the recorded wavefield, yielding various expressions for poststack and prestack migration with different recording configurations. For example, shot-record Kirchhoff migration is expressed as

$$I(\mathbf{x};\mathbf{x_s}) = \int \mathrm{d}\mathbf{x_r} \int \mathrm{d}t W \frac{\partial U(\mathbf{x_r};\mathbf{x_s};t)}{\partial t} \delta[t-(t_s+t_r)], \tag{3}$$

where $I(\mathbf{x}, \mathbf{x_s})$ is the migrated image at location $\mathbf{x}$ due to a source at $\mathbf{x_s}$, $\mathbf{x_r}$ are the receiver locations, W is a weight function, U is the recorded wavefield, δ is the Dirac delta function, and t_s and t_r are travel times from $\mathbf{x_s}$ and $\mathbf{x_r}$ to $\mathbf{x}$. The integral accumulates data samples from the receiver locations into the image at $\mathbf{x}$. Alternatively, Equation 3 allows an input sample at time $t_s + t_r$ from receiver location $\mathbf{x_r}$ to be smeared out into the isochron surface.

Kirchhoff migration is flexible, allowing the accumulation of trace values from any subvolume of the recorded traces into any subvolume of the image. This has made Kirchhoff migration very popular. However, complicated velocity models used in many depth migration projects have shown its accuracy limits, which are due mostly to a reliance on asymptotic ray theory to provide travel times and amplitudes.

Beam migration

In its simplest form, beam migration is a Kirchhoff migration with the integral over the recording surface broken into two stages: first, the domain of integration is divided into a number of overlapping regions; second, a directional decomposition of the data within each region is performed. This decomposition produces a set of local plane waves, over a range of angles, emerging within each region. Each of these local-plane wave data components is mapped back into the earth using raypaths sent in the directions of the emerging plane waves. A more complete description (Hill, 2001) uses Green's identity to downward continue the wavefields, with Gaussian beams as Green's functions. Gray et al. (2009) provide an elementary physical explanation of this complicated method, which is generally more accurate than Kirchhoff migration because it naturally allows several arrivals from each wavefield to accumulate at any image location (which is difficult for Kirchhoff migration). It has many of the flexibilities of Kirchhoff migration, but, like Kirchhoff migration, relies on asymptotic ray theory.

One-way wave-equation migration (OWEM)

Strictly speaking, applying the term "wave-equation migration" solely to migration methods that downward-continue wavefields recursively one depth step at a time is an abuse of terminology. This usage excludes Kirchhoff and beam migrations, implying that they do not derive from the wave equation, which is not true. This usage has become entrenched.

Downward continuing a wavefield from one depth to the next requires that the wavefield is propagating either up or down, but not both at once. (The wavefield can propagate laterally, but up or down is a distinguished direction.) Further, it requires an appropriate (upgoing or downgoing) wave equation. In the simple constant-velocity case, the one-way solution can be obtained easily. The two-way wave equation for wavefield P (acoustic, ignoring elastic effects such as mode conversion) is

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) P = 0. \tag{4}$$

When this equation is expressed in the frequency/horizontal wave number domain, it factors into separate equations for upgoing and downgoing wavefields. Analytical solutions of these equations provide phase shifts that perform downward continuation. Once the downward-continued wavefields have been inverse Fourier transformed from wave number back to space, they can be combined and summed over frequency using Equation 1 or 2 to produce the migrated image (Gazdag, 1978). This method, phase-shift migration, can accurately image dips up to 90°. Its only limitation is the lateral invariance of the velocity field. It has proven successful in imaging salt flanks in marine sedimentary basins. Beneath the salt flanks, where significant velocity contrast has occurred, the phase-shift image is inaccurate.

Factoring the wave equation exactly into one-way equations works only for constant velocity. When velocity varies laterally, the factoring, again with z as a distinguished direction, is approximate, incurring kinematic (time) and dynamic (amplitude) errors. A number of methods downward continue wavefields using approximate one-way wave equations. The earliest of these used finite differences to discretize the equations, using low-angle approximations (e.g., Claerbout, 1970). Later methods generalized phase-shift migration to handle lateral velocity variations, and other methods combine aspects of finite-difference and phase-shift approaches. Development of these methods continues today. Because of the errors inherent in even the best one-way wave equations, none of the methods can be completely accurate.

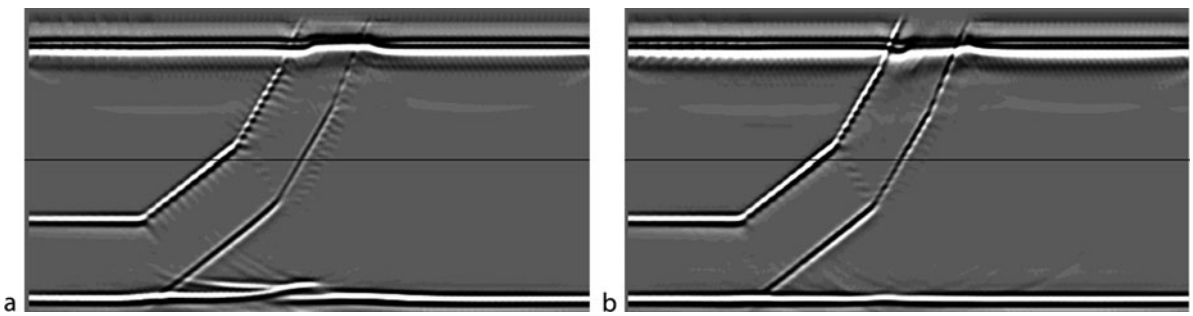
Two-way wave-equation migration (reverse-time migration)

Reverse-time migration (RTM) uses the full two-way wave equation 4. It does not operate by downward continuation of downgoing or upgoing wavefields. Instead, it takes the source and receiver wavefields at the recording surface and injects them into the earth by treating them as boundary values for the wave equation (McMechan, 1983), computing full wavefields inside the earth at each time step. The source wavefield is computed forwards in time, so successive wavefield snapshots show an expanding wavefront moving into the earth and interacting with reflectors and diffractors. The receiver wavefield is computed backwards in time (beginning with the final time sample on each trace of the recorded data), so successive snapshots show reflection events moving generally downward. Imaging is performed using a temporal version of the imaging condition (1) or (2). In principle, RTM is the most accurate migration method: it does not rely on asymptotic ray theory, and it suffers from neither the steep-dip nor the kinematic and dynamic errors of OWEM. Until recently, its use has been limited because it tends to be slower than other migration methods, and numerical approximations to the wave equation (e.g., low-order finite-difference approximations) produced errors (e.g., numerical dispersion) that other methods do not incur. However, modern computational hardware has both the speed and the memory availability to allow very good approximations to the wave equation, and high-quality RTM images are now the norm.

Migration examples

There are many reasons to migrate seismic data; we will need to be content with a few illustrations.

The first illustrates structural imaging in the presence of anisotropy. Figure 4 shows two images of a cross section designed to mimic the structural geology of a dipping thrust sheet, simplified to isolate the effect of anisotropy on image positioning and quality. The velocity is generally simple, with a constant velocity of 2,740 m/s outside the thrust sheet. The velocity complication lies in the thrust sheet, which is anisotropic (TTI), with a symmetry axis parallel to the boundaries of the sheet. In the thrust sheet, the velocity is 2,925 m/s in the direction of the symmetry axis, and higher in all other directions. Both images in Figure 4 are from prestack beam migration of the same input data set, which was generated by finite-difference elastic modeling (courtesy BP). In Figure 4a, anisotropy was ignored, and in Figure 4b, anisotropy was taken into account. The reflector below the dipping thrust sheet should be flat, as it is in Figure 4b. In Figure 4a, the flat reflector is pulled up beneath the thrust sheet, and is poorly imaged. The high-amplitude flat event cutting across the



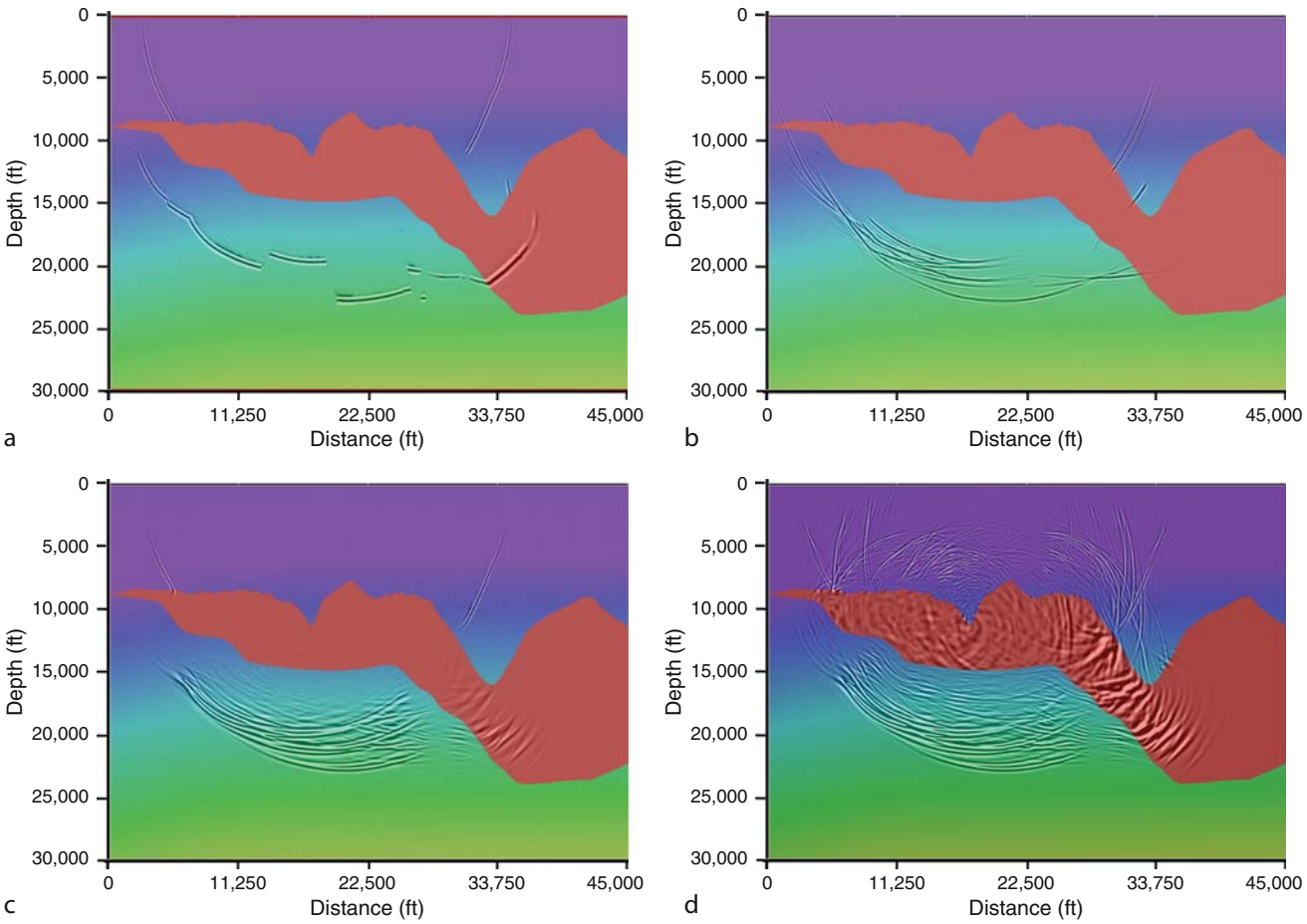
Seismic, Migration, Figure 4 Prestack depth migrated images from a synthetic thrust model data set. In (**a**), the anisotropy in the dipping thrust sheet is ignored by the migration, leading to inaccurate imaging below the thrust sheet. In (**b**), the migration has accounted for the anisotropy, leading to correct structural imaging.

entire section at shallow depths is a shear-wave artifact that has been mis-migrated by the P-wave migration program.

The second example illustrates the action of various migration methods: Kirchhoff, beam, OWEM, and RTM. Figure 5 shows four images overlaid on a geologic cross section, each obtained by migrating a single event on a single zero-offset trace through the complicated geology/velocity model shown (called Sigsbee2a). A high-velocity salt body of complex shape has been inserted into a set of lower-velocity sedimentary layers, and we note the effect of the geometry and velocity on the migration algorithms. The migrated images are, in effect, wavefronts modeled by the migration methods; full migrated images are formed by adding together such partial images from all samples on all traces. The Kirchhoff migrated image (Figure 5a) shows the effect of the assumption, made by most Kirchhoff migration programs, that at most one travel path for seismic energy connects a given upper-surface location with each subsurface location. The wavefield discontinuities in the subsalt area are physically impossible, and will cause inaccurate, noisy images in areas of structural complexity. The remaining images show wavefields of greater continuity and complexity in the subsalt area. The most complicated wavefield, produced by RTM, should produce the most accurate image. This happens in practice when the velocity field is accurate and the input traces are free from unwanted noise such as elastic-wave mode-converted events.

The evolving role of migration in seismic data processing

Before seismic processing became digital, migration was a standalone process. Pencil-and-paper or mechanical migration provided a cross section that showed where to drill. Later, poststack migration was the final seismic processing step; all other processing was intended to prepare a stacked data set for migration. With prestack migration, migration has moved to a central location in the processing flow. In tomography, it is common to process the signals on a migrated stack or in a set of CIG's. In postmigration amplitude analysis, it is common to analyze and alter the migrated waveforms. In the future, when the



Seismic, Migration, Figure 5 The action of migration, illustrated on four different migration methods: (**a**) Kirchhoff; (**b**) beam; (**c**) one-way wave-equation; (**d**) reverse-time. In each, a single event from a single zero-offset trace was migrated. The migration methods have interacted differently with the velocity model, producing different wavefronts. (From Etgen et al., 2009).

wave equation used for migration can model earth effects more completely than at present, seismic data will undergo less processing before migration, and the migrated data will predict rock properties directly.

Summary

Seismic migration has been used in prospecting for oil and natural gas since the 1920s. Since the 1970s, migration has been applied as a wave-equation process, explicitly recognizing that reflections recorded at the earth's surface are the result of wavefields propagating and reflecting inside the earth. Wavefield migration has taken several forms (time-domain, frequency-domain; finite-difference, integral) because of the variety of ways for expressing the propagation of wavefields. All these different forms of migration are still in use because each of them has features that the others lack: flexibility, steep-dip capability, etc. When computer power was limited, migration was usually performed after NMO/stack, allowing a reduction of the size of the data volume input to migration. Nowadays most migrations are performed before stack in 3-D. Another shift has been from time to depth, bringing migration closer to the problem of estimating seismic velocities inside the earth. A separate development has been the capability of migration to preserve amplitudes for estimating rock properties near reflector locations. With migration velocity estimation and amplitude analysis, migration has moved from the final step of the seismic processing flow to a more central role. This means that an increasing amount of processing and analysis is performed on migrated gathers before they are stacked to form the final structural image.

Bibliography

Berkhout, A. J., 1982. *Seismic Migration – Imaging of Acoustic Energy by Wavefield Extrapolation*. Amsterdam: Elsevier.
Bleistein, N., 1987. On the imaging of reflectors in the earth. *Geophysics*, **52**, 931–942.
Claerbout, J., 1970. Coarse-grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure. *Geophysics*, **35**, 407–418.
Claerbout, J., 1971. Toward a unified theory of reflector mapping. *Geophysics*, **36**, 467–481.
Dix, C. H., 1955. Seismic velocities from surface measurements. *Geophysics*, **20**, 68–86.
Etgen, J., Gray, S. H., and Zhang, Y., 2009. An overview of depth migration in exploration geophysics. *Geophysics*, **74**, WCA5–WCA18.
Gazdag, J., 1978. Wave equation migration with the phase-shift method. *Geophysics*, **43**, 1342–1351.
Gray, S. H., Xie, Y., Notfors, C., Zhu, T., Wang, D., and Ting, C.-O., 2009. Taking apart beam migration. *The Leading Edge*, **28**, 1098–1109.
Hagedoorn, J. G., 1954. A process of seismic reflection interpretation. *Geophysical Prospecting*, **6**, 449–453.
Hill, N. R., 2001. Prestack Gaussian-beam depth migration. *Geophysics*, **66**, 1240–1250.
Loewenthal, D., Roberson, R., Sherwood, J., and Lu, L., 1976. The wave equation applied to migration. *Geophysical Prospecting*, **24**, 380–399.
Mayne, W. H., 1962. Common reflection point horizontal data stacking techniques. *Geophysics*, **27**, 927–938.
McMechan, G. A., 1983. Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, **31**, 413–420.
Schneider, W. A., 1978. Integral formulation for migration in two and three dimensions. *Geophysics*, **43**, 49–76.
Trorey, A. W., 1970. A simple theory of seismic diffractions. *Geophysics*, **35**, 762–784.
Warner, M., 1987. Migration – why doesn't it work for deep continental data. *Geophysical Journal of the Royal Astronomical Society*, **89**, 21–26.
Yilmaz, O., 1987. *Seismic Data Processing*. Tulsa: SEG.

Cross-references

Seismic Anisotropy
Seismic Data Acquisition and Processing
Seismic Diffraction
Seismic Imaging, Overview
Seismic Properties of Rocks
Seismic Waves, Scattering
Seismic, Ray Theory
Seismic, Reflectivity Method
Seismic, Waveform Modeling and Tomography
Traveltime Tomography Using Controlled-Source Seismic Data

SEISMIC, RAY THEORY

Vlastislav Červený[1], Ivan Pšenčík[2]
[1]Department of Geophysics, Mathematics and Physics, Charles University, Praha, Czech Republic
[2]Institute of Geophysics, Academy of Sciences of Czech Republic, Praha, Czech Republic

Synonyms

Asymptotic ray theory; Ray series method; Seismic ray method

Definition

Seismic ray theory. High-frequency asymptotic method of study of seismic wavefields in complex inhomogeneous isotropic or anisotropic media with curved structural interfaces.

Introduction

The ray theory belongs to the methods most frequently used in seismology and seismic exploration for forward and inverse modeling of high-frequency seismic body waves. In smoothly varying media with smooth interfaces, it can provide useful approximate solutions of the elastodynamic equation of satisfactory accuracy. Starting from an intuitive description of the propagation of seismic waves along special trajectories - rays, it has developed into a highly sophisticated method, described briefly in this review paper.

The ray method has its advantages and disadvantages. The basic advantages are its applicability to complex,

isotropic and anisotropic, laterally varying layered media and its numerical efficiency in such computations. It provides a physical insight into the wave propagation process by separating the wavefield into individual elementary waves and by allowing their identification. In addition, it makes possible to track the paths in the medium along which energy of individual waves propagates, an aspect very important in tomography. The ray method also represents an important basis for other related, more sophisticated methods, such as the paraxial ray method, the Gaussian beam summation method, the Maslov method, the asymptotic diffraction theory, etc. The ray method also has some limitations. As mentioned above, it is approximate. It is applicable only to smooth media with smooth interfaces, in which the characteristic dimensions of inhomogeneities are considerably larger than the prevailing wavelength of the considered waves. The ray method can yield distorted results and may even fail in some special regions called *singular regions*.

The seismic ray method owes a lot to optics and radiophysics. Although the techniques used in different branches of physics are very similar, there are some substantial differences. The ray method in seismology is usually applied to more complicated structures than in optics or radiophysics. There are also different numbers and types of waves considered in different branches of physics.

The first seismological applications of ray concepts date back to the end of the 19th century. Then, only *kinematics*, specifically travel times, were used. Probably the first attempts to use also *dynamics* (amplitudes and waveforms) were made by Sir H. Jeffreys. The ray series solutions of elastodynamic equation with variable coefficients were first suggested by Babich (1956) and Karal and Keller (1959) for inhomogeneous isotropic media, and by Babich (1961) for inhomogeneous anisotropic media.

The Earth's interior is anisotropic or weakly anisotropic in some of its parts. Seismic anisotropy and its effects on wave propagation play an important role in contemporary seismology and seismic exploration. Consequently, it has also been necessary to develop the ray theory for elastic anisotropic media. It is important to emphasize that, for S waves, the ray theory for anisotropic media does not yield the ray theory for isotropic media in the zero anisotropy limit. For this reason, we describe systematically the ray theory for anisotropic media and also present corresponding formulae for isotropic media, and explain the differences between both of them.

S waves require generally a special attention. Well understood phenomenon is propagation of two separate shear waves in anisotropic media. Less understood and an underestimated phenomenon is shear-wave coupling, which occurs in weakly anisotropic media or in vicinities of shear-wave singularities. In such regions, standard ray theories for anisotropic as well as isotropic media do not work properly. Therefore, we also briefly describe the coupling ray theory for S waves, which fills the gap between ray theories for isotropic and anisotropic media.

We give here neither a detailed derivation of ray-theoretical expressions nor a relevant systematic bibliography. This would extend the text considerably. We refer, however, to several textbooks, in which the ray theory is treated in a considerably greater detail (Červený et al., 1977; Kravtsov and Orlov, 1990; Červený, 2001; Chapman, 2004). The reader may also find useful information in several review papers devoted to seismic ray theory and its various aspects (Červený et al., 1988; Virieux, 1996; Chapman, 2002; Červený et al., 2007). Examples of computations based on the ray theory can be found, for example, in Červený et al. (1977) and Gjøystdal et al. (2002). Here we refer only to papers, in which the relevant methods and procedures were first proposed, and/or which give a useful more recent treatment of the subject.

We use the following notation. We denote Cartesian coordinates x_i and time t. The dots above letters denote partial derivatives with respect to time ($\ddot{u}_i = \partial^2 u_i/\partial t^2$) and the index following the comma in the subscript indicates the partial derivative with respect to the relevant Cartesian coordinate ($u_{i,j} = \partial u_i/\partial x_j$). We consider high-frequency time-harmonic seismic body waves, with the exponential factor $\exp(-\mathrm{i}\omega t)$, where ω is fixed, positive, real-valued circular frequency. The lower-case Roman indices take the values 1, 2, 3, the upper-case indices 1, 2. Hats over bold symbols indicate 3×3 matrices, bold symbols without hats denote 2×2 matrices. The Einstein summation convention over repeating Roman indices is used, with exception of indices in parentheses.

Basic equations of the seismic ray method

For smoothly varying elastic media, the source-free *equation of motion* reads

$$\tau_{ij,j} - \rho\,\ddot{u}_i = 0. \tag{1}$$

Here $\tau_{ij}(x_n, t)$, and $u_i(x_n, t)$ are Cartesian components of stress tensor and displacement vector, respectively, and ρ is the density. In anisotropic media, the stress tensor τ_{ij} and the infinitesimal strain tensor $e_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$ are related by Hooke's law:

$$\tau_{ij} = c_{ijkl} e_{kl} = c_{ijkl} u_{k,l}. \tag{2}$$

$c_{ijkl}(x_n)$ is a tensor of elastic moduli (stiffness tensor), satisfying symmetry relations $c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}$. There are, at the most, 21 independent elastic moduli. Inserting Equation 2 into Equation 1, we get the *elastodynamic equation*

$$(c_{ijkl} u_{k,l})_{,j} - \rho\,\ddot{u}_i = 0. \tag{3}$$

In the seismic ray method, high-frequency seismic body waves propagating in smoothly varying, isotropic or anisotropic, media are studied. The formal ray series solution of the elastodynamic equation (3) for the displacement vector

$\mathbf{u}(x_n,t)$ is sought in the form of an asymptotic series in inverse powers of circular frequency ω,

$$\mathbf{u}(x_n,t) = \exp[-\mathrm{i}\omega(t-T(x_n))] \left[\mathbf{U}^{(0)}(x_n) + \frac{\mathbf{U}^{(1)}(x_n)}{(-\mathrm{i}\omega)} + \frac{\mathbf{U}^{(2)}(x_n)}{(-\mathrm{i}\omega)^2} + \ldots\right]. \quad (4)$$

Here $T(x_n)$ is the real-valued travel time, $\mathbf{U}^{(k)}$, $k = 0,1,2,\ldots$ are complex-valued *vectorial amplitude coefficients*. Surfaces $T(x_i) = const.$ are called *wavefronts* (or phase fronts). In perfectly elastic media, functions $T(x_n)$, and $\mathbf{U}^{(k)}(x_n)$ are frequency independent.

Also other forms of the ray series have been used in the seismic ray method. For example, Chapman (2004) developed the seismic ray method using the ray series for particle velocity and traction. Such a formal ray series has certain advantages with respect to Equation 4. Here, however, we consider systematically the traditional ray series (4) for the displacement vector.

Inserting Equation 4 into elastodynamic equation (3), we obtain a series in inverse powers of ω, which equals zero. Consequently, the coefficients of the individual powers of ω must also equal zero. This yields a system of equations called the basic recurrence system of equations of the ray method. This system can be used to determine the eikonal equations for travel times $T(x_n)$ and, successively the equations for the amplitude coefficients $\mathbf{U}^{(0)}(x_n)$, $\mathbf{U}^{(1)}(x_n)$, $\mathbf{U}^{(2)}(x_n)$,.... The equations for $\mathbf{U}^{(k)}(x_n)$ yield, among others, transport equations. For a detailed derivation of the basic system of equations of the ray method see Červený (2001, sect. 5.7).

The vectorial amplitude coefficients $\mathbf{U}^{(k)}(x_n)$, $k = 1,2,\ldots$, can be expressed as a sum of the *principal component* and *additional component*. The principal component of $\mathbf{U}^{(k)}(x_n)$ is the projection of $\mathbf{U}^{(k)}(x_n)$ into the unit vector parallel to the zero-order amplitude coefficient $\mathbf{U}^{(0)}(x_n)$, the additional component of $\mathbf{U}^{(k)}(x_n)$ is the remaining part of $\mathbf{U}^{(k)}(x_n)$. In this way, the additional component of the zero-order amplitude coefficient $\mathbf{U}^{(0)}(x_n)$ is zero. The complexity of the equations for higher-order amplitude coefficients $\mathbf{U}^{(k)}$ increases rapidly with increasing k. Moreover, the higher-order amplitude coefficients are inaccurate and unstable, as they are very sensitive to fine details of the medium. The instability of the amplitude coefficients increases with increasing k. For these reasons, only the zero-order coefficient $\mathbf{U}^{(0)}(x_n)$, at the most with the additional component of $\mathbf{U}^{(1)}(x_n)$, has been used in seismological applications. In the following, we shall concentrate on the zero-order ray approximation only.

The zero-order approximation of the ray method reads:

$$\mathbf{u}(x_n,t) = \mathbf{U}(x_n)\exp[-\mathrm{i}\omega(t-T(x_n))]. \quad (5)$$

In Equation 5, we have dropped the superscript (0) of $\mathbf{U}(x_n)$. We call $\mathbf{U}(x_n)$ the complex-valued *vectorial amplitude*. In smooth, laterally varying media, containing smooth structural interfaces, the zero-order approximation (5) of the ray method usually offers sufficiently accurate results, particularly for travel time $T(x_n)$. Its great advantage is that it allows one to work with frequency-independent travel time and amplitude. However, if the medium under consideration becomes more and more complex (less smooth), vectorial amplitude $\mathbf{U}(x_n)$ becomes less accurate. In structures exceeding a certain degree of complexity, the ray method may yield inaccurate results or even fail.

The first equation of the basic system of equations of the ray method reads:

$$(\Gamma_{ik} - \delta_{ik})U_k = 0, \quad i = 1,2,3. \quad (6)$$

Here $\mathbf{\Gamma}$ is the 3×3 *generalized Christoffel matrix* with elements given by the relation:

$$\Gamma_{ik} = a_{ijkl}p_j p_l. \quad (7)$$

In Equation 7, p_i are the Cartesian components of the slowness vector $\mathbf{p}$,

$$p_i = \partial T/\partial x_i \quad (8)$$

and $a_{ijkl} = c_{ijkl}/\rho$ are density-normalized elastic moduli. Note that the classical Christoffel matrix, with elements $a_{ijkl}n_j n_l$, contains components of the real-valued unit vector $\mathbf{n}$ (perpendicular to the wavefront) instead of $\mathbf{p}$. For this reason, we call $\mathbf{\Gamma}$the "generalized" Christoffel matrix. The relation between p_i and n_i is $p_i = n_i/\mathcal{C}$, where $\mathcal{C}$ is the phase velocity.

The generalized 3×3 Christoffel matrix in solid media is symmetric ($\Gamma_{ik} = \Gamma_{ki}$), positive definite ($\Gamma_{ik}a_i a_k > 0$, where a_i are components of any non-vanishing real-valued vector) and homogeneous function of the second degree in p_i ($\Gamma_{ik}(x_n, ap_j) = a^2\Gamma_{ik}(x_n,p_j)$ for any non-vanishing constant a). It has three real-valued positive eigenvalues $G_m(x_n,p_j)$, and three corresponding real-valued unit eigenvectors $\mathbf{g}^{(m)}(x_n,p_j)$, $m = 1,2,3$. G_m and $\mathbf{g}^{(m)}$ are solutions of the eigenvalue equation

$$(\Gamma_{ik} - \delta_{ik}G_m)g_k^{(m)} = 0, \quad i = 1,2,3. \quad (9)$$

Eigenvectors $\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \mathbf{g}^{(3)}$ are mutually perpendicular. Eigenvalue G_m and the relevant eigenvector $\mathbf{g}^{(m)}$ are mutually related as follows:

$$G_m = \Gamma_{ik}g_i^{(m)}g_k^{(m)} = a_{ijkl}p_j p_l g_i^{(m)}g_k^{(m)}. \quad (10)$$

For isotropic media, it is sufficient to specify elastic moduli $c_{ijkl}(x_n)$ in terms of Lamé's elastic moduli $\lambda(x_n)$ and $\mu(x_n)$, describing isotropic media, as follows:

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (11)$$

Elements of the generalized Christoffel matrix are then given by the relation:

$$\Gamma_{ik} = \frac{\lambda+\mu}{\rho}p_i p_k + \frac{\mu}{\rho}\delta_{ik}p_n p_n. \quad (12)$$

In isotropic media, the expressions for eigenvalues and eigenvectors of the generalized Christoffel matrix can be determined analytically:

$$G_1 = G_2 = \beta^2 p_k p_k, \qquad G_3 = \alpha^2 p_k p_k. \tag{13}$$

Here

$$\alpha^2 = (\lambda + 2\mu)/\rho, \qquad \beta^2 = \mu/\rho. \tag{14}$$

The eigenvector relevant to the eigenvalue G_3 equals $\mathbf{n}$, the unit vector perpendicular to the wavefront. The eigenvectors relevant to coinciding eigenvalues G_1 and G_2 are mutually perpendicular unit vectors situated arbitrarily in the plane perpendicular to $\mathbf{n}$.

Eikonal equation. Polarization vector

The comparison of the basic equation of the ray method (6) with the eigenvalue equation (9) for the 3×3 generalized Christoffel matrix shows that Equation 6 is satisfied, if the eigenvalue G_m of the generalized Christoffel matrix satisfies the relation

$$G_m(x_i, p_j) = 1, \tag{15}$$

and if the complex-valued vectorial amplitude $\mathbf{U}$ of the wave under consideration is related to eigenvector $\mathbf{g}^{(m)}$ as follows:

$$\mathbf{U} = A\mathbf{g}^{(m)}. \tag{16}$$

Equation 15 is the important *eikonal equation*. It is a nonlinear, first-order partial differential equation for travel time $T(x_n)$. Equation 16 shows that displacement vector $\mathbf{U}$ is parallel to the appropriate eigenvector $\mathbf{g}^{(m)}$. For this reason, we call $\mathbf{g}^{(m)}$ the *polarization vector*. Symbol $A(x_n)$ denotes the complex-valued, frequency-independent, *scalar amplitude*.

Taking into account that G_m is a homogeneous function of the second degree in p_i, where $\mathbf{p} = \mathcal{C}^{-1}\mathbf{n}$, we obtain $G_m(x_i, p_j) = \mathcal{C}^{-2} G_m(x_i, n_j)$. This, Equations 15 and 10 yield

$$\mathcal{C}^2(x_i, n_j) = G_m(x_i, n_j) = a_{ijkl} n_j n_l g_i^{(m)} g_k^{(m)}. \tag{17}$$

Phase velocity $\mathcal{C}$ is the velocity of the wavefront in direction $\mathbf{n}$. The phase-velocity vector $\boldsymbol{\mathcal{C}} = \mathcal{C}(x_i, n_j)\mathbf{n}$ has the direction of $\mathbf{n}$, i.e., it is perpendicular to the wavefront. It follows from Equation 17 that the squares of phase velocity $\mathcal{C}$ are eigenvalues $G_m(x_i, n_j)$ of the classical Christoffel matrix with elements $a_{ijkl} n_j n_l$.

Generally, eigenvalues G_m, $m = 1, 2, 3$, of the generalized Christoffel matrix are mutually different. They correspond to three high-frequency body waves propagating in inhomogeneous anisotropic media. We assign G_1 and G_2 to S1 and S2 waves and G_3 to P wave. If the eigenvalues are different, their polarization vectors can be determined uniquely.

If two eigenvalues coincide, we speak of the *degenerate case* of the eigenvalue problem. The corresponding eigenvectors can then be chosen as mutually perpendicular vectors situated arbitrarily in the plane perpendicular to the third eigenvector. Eigenvalues G_m may coincide *locally*, along certain lines or at certain points, which correspond to the so-called *S-wave singular directions*, or may be close to one another *globally* in a vicinity of singular directions or in *weakly anisotropic media*. The approximate but unique determination of polarization vectors in the latter situations is possible using perturbation approach (Jech and Pšenčík, 1989).

In isotropic media, the S-wave eigenvalues G_1 and G_2 coincide *globally*, see Equation 13. Consequently, in isotropic media, the S waves are controlled by a single eikonal equation and we have thus only two different eikonal equations corresponding to P and S waves. As the equations for the eigenvalues in isotropic media can be determined analytically, we can express the eikonal equations for P and S waves explicitly:

$$\alpha^2 p_k p_k = 1 \quad \text{for P waves}, \tag{18}$$

$$\beta^2 p_k p_k = 1 \quad \text{for S waves}. \tag{19}$$

In isotropic media, the generally complex-valued amplitude vector $\mathbf{U}$ can be expressed in the simple form (16) only for P waves. In this case the polarization vector $\mathbf{g}^{(3)} = \mathbf{n}$, i.e., it is perpendicular to the wavefront. For S waves, $\mathbf{U}$ must be considered in the following form:

$$\mathbf{U} = B\mathbf{g}^{(1)} + C\mathbf{g}^{(2)}. \tag{20}$$

Here $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ are two mutually perpendicular unit vectors in the plane tangent to the wavefront, i.e., perpendicular to the vector $\mathbf{n}$. The computation of $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ along the ray is explained later, see Equation 37. Symbols $B(x_n)$ and $C(x_n)$ are the corresponding, generally complex-valued scalar amplitudes.

In the seismic ray method, it is common to express the eikonal equation (15) in Hamiltonian form. Hamiltonian $\mathcal{H}(x_i, p_j)$ may be introduced in various ways. We shall consider the Hamiltonian, which is a homogeneous function of the second degree in p_i. For inhomogeneous anisotropic media, we can introduce the Hamiltonian expressed in terms of $G_m(x_i, p_j)$, see Equation 10:

$$\mathcal{H}(x_i, p_j) = \frac{1}{2} G_m(x_i, p_j) = \frac{1}{2} a_{ijkl} p_j p_l g_i^{(m)} g_k^{(m)}. \tag{21}$$

The eikonal equation (15) then yields:

$$\mathcal{H}(x_i, p_j) = \frac{1}{2}. \tag{22}$$

It holds for anisotropic as well as isotropic media.

From Equations 13 and 21, we get for isotropic inhomogeneous media:

$$\mathcal{H}(x_i, p_j) = \frac{1}{2} V^2(x_i) p_k p_k, \tag{23}$$

where $V = \alpha$ for P waves and $V = \beta$ for S waves.

Ray tracing and travel-time computation

The eikonal equation in Hamiltonian form (22), with $p_j = \partial T/\partial x_j$, is a non-linear partial differential equation of the first order for travel time $T(x_i)$. It can be solved by the method of characteristics. The characteristics of eikonal equation (22) are spatial trajectories, along which Equation 22 is satisfied, and along which travel time T can be computed by quadratures. The characteristics of the eikonal equation represent *rays*.

The characteristics of the eikonal equation expressed in general Hamiltonian form are described by a system of non-linear, ordinary differential equations of the first order:

$$\frac{\mathrm{d}x_i}{\mathrm{d}u} = \frac{\partial \mathcal{H}}{\partial p_i}, \quad \frac{\mathrm{d}p_i}{\mathrm{d}u} = -\frac{\partial \mathcal{H}}{\partial x_i}, \quad \frac{\mathrm{d}T}{\mathrm{d}u} = p_k \frac{\partial \mathcal{H}}{\partial p_k}. \tag{24}$$

Here u is a real-valued parameter along the ray. The relation between parameter u and the travel time along the ray depends on the form of the Hamiltonian used, see the last equation in Equations 24. For Hamiltonians, which are homogeneous functions of the second degree in p_i, the Euler equation for homogeneous functions yields $p_k \partial \mathcal{H}/\partial p_k = 2\mathcal{H}$. If we consider Hamiltonian (21), we get $\mathrm{d}T/\mathrm{d}u = 1$ from Equations 24. For travel time T along the ray, denoted $\tau = T$, Equations 24 simplify to:

$$\frac{\mathrm{d}x_i}{\mathrm{d}\tau} = \frac{\partial \mathcal{H}}{\partial p_i}, \quad \frac{\mathrm{d}p_i}{\mathrm{d}\tau} = -\frac{\partial \mathcal{H}}{\partial x_i}. \tag{25}$$

This system of equations is usually called the *ray tracing system*. Solution of the ray tracing system (25) with appropriate initial conditions yields $x_i(\tau)$, the coordinates of points along the ray trajectory, and $p_i(\tau)$, the Cartesian components of the slowness vectors along the ray. The travel time T along the ray is obtained automatically, $T = \tau$.

Inserting Equation 21 in Equations 25, we obtain the ray tracing system for m-th wave in inhomogeneous anisotropic media:

$$\begin{aligned} \frac{\mathrm{d}x_i}{\mathrm{d}\tau} &= a_{ijkl} p_l g_j^{(m)} g_k^{(m)}, \\ \frac{\mathrm{d}p_i}{\mathrm{d}\tau} &= -\frac{1}{2} \frac{\partial a_{jkln}}{\partial x_i} p_k p_n g_j^{(m)} g_l^{(m)}. \end{aligned} \tag{26}$$

In the derivation of the first set of Equations 26 for $\partial \mathcal{H}/\partial p_i$, we took into account that $\Gamma_{ik} \partial (g_i^{(m)} g_k^{(m)})/\partial p_n = 0$. An alternative version of ray tracing equation (26) was derived by Červený (1972), in which the eigenvectors $\mathbf{g}^{(m)}$ are not used.

The initial conditions for the ray tracing system (26) are $x_i = x_{0i}$, $p_i = p_{0i}$, where x_{0i} and p_{0i} satisfy the eikonal equation (22), corresponding to the wave we wish to compute (P, S1 or S2). Components p_{0i} of the initial slowness vector $\mathbf{p}_0$ can be then expressed as $p_{0i} = n_{0i}/\mathcal{C}(x_{0i})$, where $\mathcal{C}$ is the relevant phase velocity. The eikonal equation (22) is then satisfied along the whole ray.

In inhomogeneous isotropic media, the ray tracing system (25) with Equation 23 yields

$$\frac{\mathrm{d}x_i}{\mathrm{d}\tau} = V^2 p_i, \quad \frac{\mathrm{d}p_i}{\mathrm{d}\tau} = -\frac{\partial \ln V}{\partial x_i}. \tag{27}$$

The initial conditions for the ray tracing system (27) are again $x_i = x_{0i}$, $p_i = p_{0i}$, where $p_{0i} = n_{0i}/V(x_{0i})$. Here $V = \alpha$ for P waves, and $V = \beta$ for S waves.

As τ is the travel time along the ray, $\mathrm{d}x_i/\mathrm{d}\tau$ represent the Cartesian components $\mathcal{U}_i$ of the *ray-velocity vector* $\boldsymbol{\mathcal{U}}$ of the m-th wave:

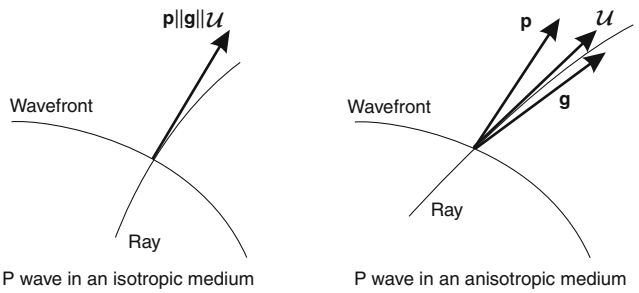
$$\mathcal{U}_i = a_{ijkl} p_l g_j^{(m)} g_k^{(m)}. \tag{28}$$

In non-dissipative anisotropic media, the ray-velocity vector $\boldsymbol{\mathcal{U}}$ is also called the group-velocity vector or the energy-velocity vector. As indicated by the name, the energy velocity vector $\boldsymbol{\mathcal{U}}$ represents the velocity of the energy propagation.

In anisotropic media, the ray-velocity vector $\boldsymbol{\mathcal{U}}$ must be strictly distinguished from the phase-velocity vector $\boldsymbol{\mathcal{C}}$. In inhomogeneous anisotropic media, the ray-velocity and phase-velocity vectors $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{C}}$ are generally different, both in size and direction. Vector $\boldsymbol{\mathcal{U}}$ is always greater than $\boldsymbol{\mathcal{C}}$. The two vectors are equal (in size and direction) only in special directions, called *longitudinal directions*.

In inhomogeneous isotropic media, Equation 28 for the ray-velocity vector yields $\boldsymbol{\mathcal{U}} = V^2 \mathbf{p}$. For the phase-velocity vector, using Equation 17, we get $\boldsymbol{\mathcal{C}} = V^2 \mathbf{p}$. In both cases, $V = \alpha$ for P waves, and $V = \beta$ for S waves. Thus, the ray-velocity and phase-velocity vectors are identical in isotropic media.

Figure 1 shows mutual orientation of ray-velocity vector $\boldsymbol{\mathcal{U}}$, phase-velocity vector $\boldsymbol{\mathcal{C}}$ (parallel to slowness vector $\mathbf{p}$) and polarization vector $\mathbf{g}$ of a P wave propagating in an isotropic (left) and anisotropic (right) medium. While $\boldsymbol{\mathcal{U}}$, $\boldsymbol{\mathcal{C}}$ and $\mathbf{g}$ are parallel in isotropic media, they generally differ in anisotropic media. For S waves, the vectors $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{C}}$ have similar orientation as in the case of P waves. The polarization vectors $\mathbf{g}$ are, however, perpendicular



Seismic, Ray Theory, Figure 1 Slowness vector $\mathbf{p}$ (perpendicular to the wavefront), ray-velocity vector $\boldsymbol{\mathcal{U}}$ (tangent to the ray) and polarization vector $\mathbf{g}$ of a P wave propagating in an isotropic (*left*) and anisotropic (*right*) medium. For simplicity, the three vectors in the right-hand plot are shown in one plane. In general, this is not the case in anisotropic media.

(isotropic medium) or nearly perpendicular (anisotropic medium) to the ray.

Ray tracing systems (26) and (27) can be simply solved if the initial values x_{0i} and p_{0i} are specified at some point S. We then speak of *initial-value ray tracing*. The standard numerical procedures of solving the system of ordinary differential equations of the first order with specified initial conditions can then be used (Runge-Kutta, etc.). A very important role in seismology is played by *boundary-value ray tracing*, in which we seek the ray, satisfying some boundary conditions. The typical boundary-value problem is *two-point ray tracing*, in which we seek the ray connecting two specified points. Mostly, the controlled initial-value ray tracing (controlled shooting method) is used to solve this problem (Červený et al., 2007). Boundary-value ray tracing is considerably more complicated than initial-value ray tracing.

There are four important differences between initial-value ray tracing in isotropic and anisotropic media. First: In anisotropic media, we deal with three waves, P, S1 and S2, in isotropic media with two waves, P and S, only. Second: In inhomogeneous anisotropic media, ray tracing system (26) is the same for all three waves. The wave under consideration is specified by the initial conditions, which must satisfy the eikonal equation of the considered wave. In isotropic inhomogeneous media, the ray tracing systems are different for P and S waves, see Equations 27 with $V = \alpha$ and $V = \beta$, respectively. Third: In isotropic media, the initial direction of the slowness vector specifies directly the initial direction of the ray (as the tangent to the ray and the slowness vector have the same directions). In anisotropic media, the direction of the ray is, generally, different from the direction of the slowness vector. Nevertheless, we have to use p_{0i} as the initial values for the ray tracing system. The ray-velocity vector $\boldsymbol{\mathcal{U}}$ can be simply calculated from slowness vector $\mathbf{p}$ at any point of the ray, including the initial point. Fourth: Ray tracing for P and S waves is regular everywhere in inhomogeneous isotropic media. In anisotropic media, problems arise with tracing S-wave rays in vicinities of singular directions, or if medium is nearly isotropic (quasi-isotropic).

The problem of ray tracing and travel-time computation in inhomogeneous media has been broadly discussed in the seismological literature; particularly for inhomogeneous isotropic media. Many ray tracing systems and many suitable numerical procedures for performing ray tracing have been proposed. For 1-D isotropic media (vertically inhomogeneous, radially symmetric), the ray tracing systems may be simplified so that they reduce to simple quadratures, well known from classical seismological textbooks (Aki and Richards, 1980). Standard programs for ray tracing and travel-time computations in laterally varying isotropic and anisotropic structures are available, see, for example, program packages SEIS (2D isotropic models), CRT and ANRAY (3D isotropic/anisotropic models) at http://sw3d.cz/. Programs for anisotropic media have, however, problems with S-wave computations in quasi-isotropic media and in the vicinities of shear-wave singularities. In such cases, the standard ray theory should be replaced by the coupling ray theory. Numerical procedures based on the coupling ray theory are, unfortunately, rare.

Ray tracing may also serve as a basis for the so-called *wavefront construction method* (Gjøystdal et al., 2002). In this case, for a selected wave, wavefronts with travel times $T = T_0 + k\Delta T$ are computed successively from the previous wavefronts with travel times $T = T_0 + (k-1)\Delta T$. The wavefront construction method has found broad applications in seismic exploration.

Let us consider a two-parametric system of rays, call it the *ray field*, and specify the individual rays in the ray field by *ray parameters* γ_1, γ_2. Ray parameters γ_1, γ_2 may represent, e.g., the take-off angles at a point source, or the curvilinear Gaussian coordinates of initial ray points along the initial surface. The family of rays with ray parameters within the limit $[\gamma_1, \gamma_1 + \mathrm{d}\gamma_1]$, $[\gamma_2, \gamma_2 + \mathrm{d}\gamma_2]$, is called the *elementary ray tube* or briefly the ray tube. We further introduce *ray coordinates* $\gamma_1, \gamma_2, \gamma_3$ in such a way that γ_1, γ_2 are ray parameters, and γ_3 is some monotonic parameter along a ray (arclength s, travel time τ, etc.). Here we consider $\gamma_3 = \tau$, but our results may be simply modified for any other monotonic parameter γ_3. We further introduce the 3×3 transformation matrix $\hat{\mathbf{Q}}$ from ray to Cartesian coordinates with elements $Q_{ij} = \partial x_i / \partial \gamma_j$. The Jacobian of transformation from ray to Cartesian coordinates, $\det \hat{\mathbf{Q}}$, can be expressed as follows:

$$\det \hat{\mathbf{Q}}(\tau) = (\partial \mathbf{x}(\tau)/\partial \gamma_1 \times \partial \mathbf{x}(\tau)/\partial \gamma_2)^T \boldsymbol{\mathcal{U}}(\tau). \qquad (29)$$

The vectorial product in Equation 29 has the direction of the normal to the wavefront, specified by $\mathbf{n} = \mathcal{C}\,\mathbf{p}$. As $\mathbf{p}(\tau) \cdot \boldsymbol{\mathcal{U}}(\tau) = 1$, see Equations 28, 10, and 15, we also obtain

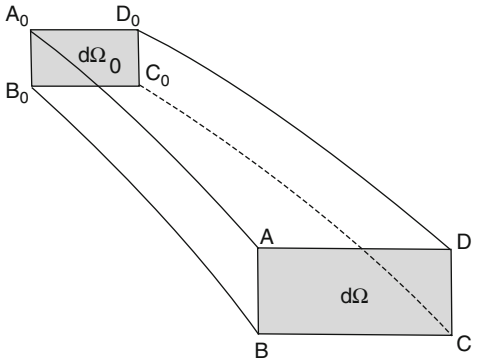
$$\det \hat{\mathbf{Q}}(\tau) = \pm \mathcal{C}(\tau) |(\partial \mathbf{x}(\tau)/\partial \gamma_1 \times \partial \mathbf{x}(\tau)/\partial \gamma_2)|. \qquad (30)$$

Thus Jacobian $\det \hat{\mathbf{Q}}(\tau)$ equals $\pm\, \mathcal{C}(\tau) \mathrm{d}\Omega(\tau)$, where $\mathrm{d}\Omega(\tau) = |(\partial \mathbf{x}(\tau)/\partial \gamma_1 \times \partial \mathbf{x}(\tau)/\partial \gamma_2)|$ is the scalar surface element cut out of the wavefront by the ray tube. It measures the expansion or contraction of the ray tube, see Figure 2. For this reason, the 3×3 matrix $\hat{\mathbf{Q}}(\tau)$ is also often called the *geometrical spreading matrix* and various quantities related to $\det \hat{\mathbf{Q}}(\tau)$ are called *geometrical spreading*. It plays an important role in the computation of the ray-theory amplitudes.

Transport equation. Computation of ray-theory amplitudes

The second equation of the basic system of equations of the ray method yields the transport equation for the scalar ray-theory amplitude $A(x_i)$. The transport equation is a partial differential equation of the first order. It can be expressed in several forms. One of them, valid both for isotropic and anisotropic media, reads

$$\nabla \cdot (\rho A^2 \boldsymbol{\mathcal{U}}) = 0. \qquad (31)$$



Seismic, Ray Theory, Figure 2 Elementary ray tube. $d\Omega_0$ and $d\Omega$ are scalar surface elements cut out of the wavefront by the ray tube. This means that in isotropic media, the normals to $d\Omega_0$ and $d\Omega$ are parallel to rays. In anisotropic media, they are not.

It is common to solve the transport equation along the ray. $\nabla \cdot \boldsymbol{\mathcal{U}}$ can then be expressed as follows:

$$\nabla \cdot \boldsymbol{\mathcal{U}} = \mathrm{d}[\ln(\det \hat{\mathbf{Q}})]/\mathrm{d}\tau \tag{32}$$

(Červený, 2001, Equation 3.10.24). Inserting Equation 32 into Equation 31 yields the transport equation in the form of the first-order ordinary differential equation along the ray:

$$\mathrm{d}\big(\rho(\tau)A^2(\tau)\det \hat{\mathbf{Q}}(\tau)\big)/\mathrm{d}\tau = 0. \tag{33}$$

This yields a simple form of the continuation relation for $A(\tau)$ along the ray:

$$A(\tau) = \left[\frac{\rho(\tau_0)\det \hat{\mathbf{Q}}(\tau_0)}{\rho(\tau)\det \hat{\mathbf{Q}}(\tau)}\right]^{1/2} A(\tau_0). \tag{34}$$

We obtain another suitable continuation relation for amplitudes along the ray by introducing a special local Cartesian coordinate system y_1, y_2, y_3, varying along the ray. We call it the *wavefront orthonormal coordinate system*. At any point of the ray specified by $\gamma_3 = \tau$, the y_3 axis is parallel to slowness vector $\mathbf{p}$, and the y_1, y_2 axes are confined to the plane tangential to the wavefront at $\gamma_3 = \tau$. Axes y_1 and y_2 are mutually perpendicular. If we denote the 3×3 transformation matrix from ray coordinates to wavefront orthonormal coordinates by $\hat{\mathbf{Q}}^{(y)}$, then

$$\det \hat{\mathbf{Q}}(\tau) = \det \hat{\mathbf{Q}}^{(y)}(\tau) = \mathcal{C}(\tau)\det \mathbf{Q}^{(y)}(\tau). \tag{35}$$

Here $\mathcal{C}(\tau)$ is the phase velocity, and $\mathbf{Q}^{(y)}(\tau)$ is the 2×2 upper-left submatrix of $\hat{\mathbf{Q}}^{(y)}(\tau)$. Using Equation 35 in Equation 34, we obtain the continuation relation in an alternative form:

$$A(\tau) = \left[\frac{\rho(\tau_0)\mathcal{C}(\tau_0)\det \mathbf{Q}^{(y)}(\tau_0)}{\rho(\tau)\mathcal{C}(\tau)\det \mathbf{Q}^{(y)}(\tau)}\right]^{1/2} A(\tau_0). \tag{36}$$

An important property of continuation relation (36) is that $\det \mathbf{Q}^{(y)}(\tau)$ is uniquely determined by coordinates y_1 and y_2, confined to the plane tangential to the wavefront at τ. Thus, Equation 36 remains valid for any coordinate systems q_i (even nonorthogonal), in which mutually perpendicular coordinate axes q_1 and q_2 are confined to the plane tangential to the wavefront, but the axis q_3 is taken in a different way than y_3, for example along the ray. This is, e.g., the case of the well-known *ray-centered coordinate system* q_1, q_2, q_3. We have $\det \mathbf{Q}^{(q)}(\tau) = \det \mathbf{Q}^{(y)}(\tau)$.

Transport equations for P and S waves in isotropic media may be also expressed in the form of Equation 31. The expression is straightforward for P waves. For S waves, transport equations for scalar amplitudes B and C in Equation 20 are generally coupled. They decouple only if the unit vectors $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ in Equation 20 satisfy the following relation along the ray:

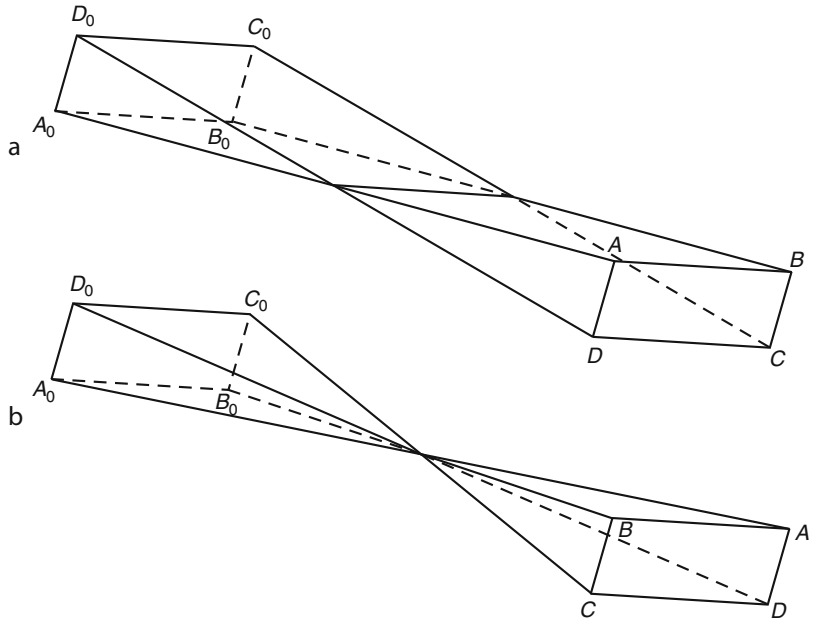
$$\mathrm{d}\mathbf{g}^{(M)}/\mathrm{d}\tau = (\mathbf{g}^{(M)} \cdot \nabla\beta)\mathbf{n}, \quad M = 1, 2. \tag{37}$$

In the terminology of the Riemanian geometry, vector $\mathbf{g}^{(M)}$ satisfying Equation 37 is transported parallelly along the ray. If $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ are chosen as mutually perpendicular and perpendicular to $\mathbf{n}$ at one point of the ray, Equation 37 guarantees that they have these properties at any point of the ray. Consequently, $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ are always perpendicular to the ray and do not rotate around it as the S wave progresses. As $\mathbf{g}^{(1)}$, $\mathbf{g}^{(2)}$ and $\mathbf{n}$ are always orthonormal, and $\mathbf{n}$ is known at any point of the ray, it is not necessary to use Equation 37 to compute both vectors $\mathbf{g}^{(M)}$. One of them can be determined from the orthonormality condition, once the other has been computed using Equation 37.

Quantity $\det \hat{\mathbf{Q}}(\tau)$ in Equation 34 may be zero at some point $\tau = \tau^C$. This means that the cross-sectional area of the ray tube shrinks to zero at $\tau = \tau^C$. The relevant point $\tau = \tau^C$ of the ray is called the *caustic point*. At the caustic point, the ray solution is singular and yields an infinite amplitude there. In passing through the caustic point τ^C along the ray, the argument of $[\det \hat{\mathbf{Q}}(\tau)]^{1/2}$ may change by $\pm\pi/2$ or $\pm\pi$ (Kravtsov and Orlov, 1999). The former case corresponds to the caustic point of the first order, see Figure 3a, during which the ray tube shrinks to an elementary arc, the latter case corresponds to the caustic point of the second order, see Figure 3b, during which the ray tube shrinks to a point. It is common to introduce the *phase shift due to caustic* $T^C(\tau, \tau_0)$ using the relation

$$\left[\frac{\det \hat{\mathbf{Q}}(\tau_0)}{\det \hat{\mathbf{Q}}(\tau)}\right]^{1/2} = \left|\frac{\det \hat{\mathbf{Q}}(\tau_0)}{\det \hat{\mathbf{Q}}(\tau)}\right|^{1/2} \exp[\mathrm{i}T^C(\tau, \tau_0)] \tag{38}$$

if caustic point τ^C is situated between τ_0 and τ. The phase shift due to the caustic is cumulative. If the ray passes through several caustic points along the ray between τ_0 and τ, the phase shift due to caustics is the sum of the individual phase shifts. It is often expressed in the form $T^C(\tau, \tau_0) = -\frac{1}{2}\pi k(\tau, \tau_0)$, where $k(\tau, \tau_0)$ is an integer,



Seismic, Ray Theory, Figure 3 Caustic points of (**a**) the first order and (**b**) second order. (Figure 3.13 of Červený, 2001.)

called the KMAH index (to acknowledge the work by - Keller, Maslov, Arnold and Hörmander in this field). The continuation relation for ray-theory amplitudes (34) can then be modified to read:

$$A(\tau) = \left(\frac{\rho(\tau_0)|\det \hat{\mathbf{Q}}(\tau_0)|}{\rho(\tau)|\det \hat{\mathbf{Q}}(\tau)|} \right)^{1/2} \exp[\mathrm{i}T^C(\tau, \tau_0)] A(\tau_0). \tag{39}$$

Equation 36 can be transformed to the analogous form as Equation 39 as the zeros of $\det \mathbf{Q}^{(y)}(\tau)$ are situated at the same points τ^C on the ray as the zeros of $\det \hat{\mathbf{Q}}(\tau)$.

The KMAH index can be calculated along the ray as a byproduct of dynamic ray tracing. For detailed derivations and discussion see Bakker (1998) and Klimeš (2010).

There are some differences between the KMAH indices along the rays in isotropic and anisotropic media. In isotropic media, the KMAH index always increases when the ray passes through a new caustic point, either by one or two. In anisotropic media, however, it may also decrease by one or two at some caustic points. This happens only for S waves as a consequence of the concave form of the slowness surface of the corresponding S wave.

Dynamic ray tracing. Paraxial approximations

As we can see in Equation 34, the computation of the ray-theory amplitudes requires knowledge of $\det \hat{\mathbf{Q}}$, where $\hat{\mathbf{Q}}(\tau)$ characterizes the properties of the ray field in the vicinity of the ray under consideration. $\hat{\mathbf{Q}}(\tau)$ can be computed by the procedure called *dynamic* (or paraxial) *ray tracing*. In addition to $\hat{\mathbf{Q}}(\tau)$ with elements $Q_{ij}(\tau) = \partial x_i / \partial \gamma_j$, we also have to introduce a new 3×3 matrix $\hat{\mathbf{P}}(\tau)$ with elements $P_{ij}(\tau) = \partial p_i / \partial \gamma_j$. The equation for P_{ij} must be included to obtain the linear dynamic ray tracing system. Differentiating ray tracing equations (25) with respect to γ_j, we can easily obtain a system of linear ordinary differential equations of the first order for Q_{ij} and P_{ij},

$$\frac{\mathrm{d}Q_{ij}}{\mathrm{d}\tau} = \frac{\partial^2 \mathcal{H}}{\partial p_i \partial x_k} Q_{kj} + \frac{\partial^2 \mathcal{H}}{\partial p_i \partial p_k} P_{kj},$$

$$\frac{\mathrm{d}P_{ij}}{\mathrm{d}\tau} = -\frac{\partial^2 \mathcal{H}}{\partial x_i \partial x_k} Q_{kj} - \frac{\partial^2 \mathcal{H}}{\partial x_i \partial p_k} P_{kj}, \tag{40}$$

see Červený (1972). This system is usually called the *dynamic ray tracing system*, and the relevant procedure dynamic ray tracing. It can be solved along a given ray Ω, or together with it.

The dynamic ray tracing system (40) may be expressed in various forms. Instead of Cartesian coordinates x_i, we can use the wavefront orthonormal coordinates y_i, or the ray-centered coordinates q_i. Then, instead of the 3×3 matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$, it is sufficient to seek the 2×2 matrices $\mathbf{Q}^{(y)}$, $\mathbf{P}^{(y)}$ or $\mathbf{Q}^{(q)}$, $\mathbf{P}^{(q)}$. This reduces the number of DRT equations, but complicates their right-hand sides (Červený 2001, sect. 4.2).

As the dynamic ray tracing system (40) is of the first order and linear, we can compute its fundamental matrix consisting of six linearly independent solutions. The 6×6 fundamental matrix of system (40) specified by the 6×6 identity matrix at an arbitrary point $\tau = \tau_0$ of the ray is called the *ray propagator matrix* and denoted by $\mathbf{\Pi}(\tau, \tau_0)$.

The 6×6 ray propagator matrix $\mathbf{\Pi}(\tau, \tau_0)$ is symplectic:

$$\mathbf{\Pi}^T(\tau, \tau_0)\mathbf{J}\mathbf{\Pi}(\tau, \tau_0) = \mathbf{J}, \quad \text{with } \mathbf{J} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix} \tag{41}$$

If we know the matrices $\hat{\mathbf{Q}}(\tau_0)$, $\hat{\mathbf{P}}(\tau_0)$, we can compute $\hat{\mathbf{Q}}(\tau)$, $\hat{\mathbf{P}}(\tau)$ at any point τ of the ray by a simple matrix multiplication

$$\begin{pmatrix} \hat{\mathbf{Q}}(\tau) \\ \hat{\mathbf{P}}(\tau) \end{pmatrix} = \mathbf{\Pi}(\tau, \tau_0) \begin{pmatrix} \hat{\mathbf{Q}}(\tau_0) \\ \hat{\mathbf{P}}(\tau_0) \end{pmatrix}. \tag{42}$$

The ray propagator matrix $\mathbf{\Pi}(\tau, \tau_0)$ satisfies the chain rule, $\mathbf{\Pi}(\tau, \tau_0) = \mathbf{\Pi}(\tau, \tau_1)\mathbf{\Pi}(\tau_1, \tau_0)$, where point τ_1 is situated arbitrarily on the ray. It is simple to compute the inverse of $\mathbf{\Pi}(\tau, \tau_0)$: $\mathbf{\Pi}^{-1}(\tau, \tau_0) = \mathbf{\Pi}(\tau_0, \tau)$. We can express $\mathbf{\Pi}(\tau, \tau_0)$ in the following way:

$$\mathbf{\Pi}(\tau, \tau_0) = \begin{pmatrix} \hat{\mathbf{Q}}_1(\tau, \tau_0) & \hat{\mathbf{Q}}_2(\tau, \tau_0) \\ \hat{\mathbf{P}}_1(\tau, \tau_0) & \hat{\mathbf{P}}_2(\tau, \tau_0) \end{pmatrix}, \tag{43}$$

where $\hat{\mathbf{Q}}_1(\tau, \tau_0), \hat{\mathbf{Q}}_2(\tau, \tau_0), \hat{\mathbf{P}}_1(\tau, \tau_0)$ and $\hat{\mathbf{P}}_2(\tau, \tau_0)$ are 3×3 matrices.

Equation 42 can be used to obtain a very important quantity – the 3×3 matrix $\hat{\mathbf{M}}(\tau)$ of second derivatives of the travel-time field with respect to Cartesian coordinates, with elements $M_{ij} = \partial^2 T / \partial x_i \partial x_j$:

$$\hat{\mathbf{M}}(\tau) = \hat{\mathbf{P}}(\tau)(\hat{\mathbf{Q}}(\tau))^{-1}. \tag{44}$$

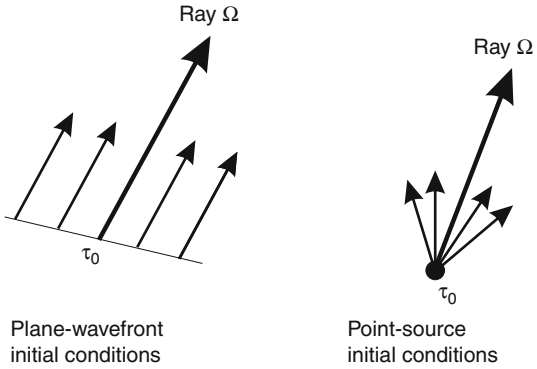
Matrix $\hat{\mathbf{M}}(\tau)$ plays an important role in the computation of travel time not only along the ray Ω, but also in its "quadratic" paraxial vicinity:

$$T(\mathbf{x}) = T(\mathbf{x}^\Omega) + (\mathbf{x} - \mathbf{x}^\Omega)^T \mathbf{p}(\tau) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^\Omega)^T \hat{\mathbf{M}}(\tau)(\mathbf{x} - \mathbf{x}^\Omega). \tag{45}$$

In Equation 45, $\mathbf{x}$ denotes an arbitrary point in the paraxial vicinity of the ray Ω, close to point $\mathbf{x}^\Omega = \mathbf{x}^\Omega(\tau)$ on the ray Ω; slowness vector $\mathbf{p}(\tau)$ and the matrix $\hat{\mathbf{M}}(\tau)$ are given at $\mathbf{x}^\Omega$. The possibility of computing the travel time in the paraxial vicinity of the ray has many important applications.

The properties of the 6×6 ray propagator matrix $\mathbf{\Pi}(\tau, \tau_0)$ described above remain valid even for the 4×4 ray propagator matrices $\mathbf{\Pi}^{(y)}(\tau, \tau_0)$ or $\mathbf{\Pi}^{(q)}(\tau, \tau_0)$ expressed in wavefront orthonormal coordinates y_i or ray-centered coordinates q_i. The ray propagator matrices $\mathbf{\Pi}^{(y)}(\tau, \tau_0)$ and $\mathbf{\Pi}^{(q)}(\tau, \tau_0)$ are identical, therefore, they can be expressed in terms of the same 2×2 matrices $\mathbf{Q}_1(\tau, \tau_0)$, $\mathbf{Q}_2(\tau, \tau_0)$, $\mathbf{P}_1(\tau, \tau_0)$ and $\mathbf{P}_2(\tau, \tau_0)$. Matrices $\mathbf{Q}_1(\tau, \tau_0), \mathbf{P}_1(\tau, \tau_0)$ correspond to the plane-wavefront initial conditions at τ_0, and matrices $\mathbf{Q}_2(\tau, \tau_0), \mathbf{P}_2(\tau, \tau_0)$ to the point-source initial conditions at τ_0, see Figure 4. The 2×2 matrix $\mathbf{Q}_2(\tau, \tau_0)$ plays an important role in computing the ray-theory Green function. The quantity

$$\mathcal{L}(\tau, \tau_0) = |\det \mathbf{Q}_2(\tau, \tau_0)|^{1/2} \tag{46}$$



Seismic, Ray Theory, Figure 4 Plane-wavefront and point-source initial conditions for dynamic ray tracing. In anisotropic media, rays are not perpendicular to the wavefront.

is called the *relative geometrical spreading*. It corresponds to a point source.

As in Equation 44, we can define the 2×2 matrix of the second derivatives of the travel-time field with respect to y_1, y_2 or q_1, q_2 as follows:

$$\mathbf{M}(\tau) = \mathbf{P}(\tau)(\mathbf{Q}(\tau))^{-1}. \tag{47}$$

We will now briefly summarize several useful ray-theory quantities and applications, which rely fully or partly on dynamic ray tracing. For derivations and more detailed explanations, see Červený (2001, Chap. 4), where also many other applications and references can be found: (1) Paraxial travel times. (2) Paraxial slowness vectors. (3) Paraxial rays. (4) Curvature of the wavefront. (5) Matrix of geometrical spreading $\hat{\mathbf{Q}}$ and the relevant matrix $\hat{\mathbf{P}}$. (6) Continuation relations for ray-theory amplitudes along the ray. (7) Relative geometrical spreading. (8) Phase shift due to caustics. (9) Ray-theory elastodynamic Green function. (10) Higher-order spatial derivatives of the travel-time field. (11) Fresnel volumes and Fresnel zones. (12) Surface-to-surface propagator matrix. (13) Boundary-value problems in four-parametric system of paraxial rays, including two-point ray tracing. (14) Factorization of the geometrical spreading.

Dynamic ray tracing is also needed in the investigation of ray chaos and in computations of Lyapunov exponents, in the ray-perturbation methods and in modifications and extensions of the ray method such as Maslov method, Gaussian beam and Gaussian packet summation methods, in Kirchhoff-Helmholtz method and in various diffraction methods.

Coupling ray theory for S waves in anisotropic media

In inhomogeneous weakly anisotropic media, the standard ray theory described above yields distorted results since it is unable to describe the coupling of S1 and S2 waves propagating with approximately equal phase velocities. This problem can be removed by using the coupling ray

theory. In the coupling ray theory, the amplitudes of the two S waves can be computed along a trajectory called the *common ray* (Bakker, 2002; Klimeš, 2006). The closer the common ray approximates actual S-wave rays, the more accurate results the coupling ray theory yields. The common rays can be constructed in a reference isotropic medium or in the actual anisotropic medium. A convenient option is to compute common rays using ray tracing equations (25) with the Hamiltonian given as

$$\mathcal{H}(x_i, p_j) = \frac{1}{4}[G_1(x_i, p_j) + G_2(x_i, p_j)]. \tag{48}$$

In Equation 48, G_1 and G_2 are eigenvalues of the Christoffel matrix $\mathbf{\Gamma}$, see equation (7), corresponding to S1 and S2 waves.

The coupling ray theory solution is sought in the form (Coates and Chapman, 1990; Bulant and Klimeš, 2002):

$$\begin{aligned}\mathbf{u}(\tau, t) = A(\tau)[r_1(\tau)\mathbf{g}^{(1)}(\tau)\exp(\mathrm{i}\omega\tau_1) \\ + r_2(\tau)\mathbf{g}^{(2)}(\tau)exp(\mathrm{i}\omega\tau_2)]\exp(-\mathrm{i}\omega t).\end{aligned} \tag{49}$$

Here, $A(\tau)$ is the scalar amplitude (34) or (36) calculated along the common ray. The symbols $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ denote the S-wave eigenvectors of the generalized Christoffel matrix $\Gamma(x_i, p_j)$ calculated along the common ray. The travel times τ_1 and τ_2 are travel times corresponding to the above vectors $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$. They can be obtained by quadratures along the common ray:

$$\begin{aligned}\mathrm{d}\tau_1/\mathrm{d}\tau &= [\Gamma_{ik} g_i^{(1)} g_k^{(1)}]^{-1/2}, \\ \mathrm{d}\tau_2/\mathrm{d}\tau &= [\Gamma_{ik} g_i^{(2)} g_k^{(2)}]^{-1/2}.\end{aligned} \tag{50}$$

The amplitude factors r_1 and r_2 are solutions of two coupled ordinary differential equations (Coates and Chapman, 1990):

$$\begin{pmatrix}\mathrm{d}r_1/\mathrm{d}\tau \\ \mathrm{d}r_2/\mathrm{d}\tau\end{pmatrix} = \frac{\mathrm{d}\varphi}{\mathrm{d}\tau}\begin{pmatrix}0 & \exp(\mathrm{i}\omega[\tau_2(\tau)-\tau_1(\tau)]) \\ -\exp(\mathrm{i}\omega[\tau_1(\tau)-\tau_2(\tau)]) & 0\end{pmatrix}\begin{pmatrix}r_1 \\ r_2\end{pmatrix}, \tag{51}$$

where the angular velocity $\mathrm{d}\varphi/\mathrm{d}\tau$ of the rotation of the eigenvectors $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ is given by

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = \mathbf{g}^{(2)}\frac{\mathrm{d}\mathbf{g}^{(1)}}{\mathrm{d}\tau} = -\mathbf{g}^{(1)}\frac{\mathrm{d}\mathbf{g}^{(2)}}{\mathrm{d}\tau}. \tag{52}$$

For detailed description of the algorithm, see Bulant and Klimeš (2002).

There are many possible modifications and approximations of the coupling ray theory. In some of them, the amplitude vector $\mathbf{U}$ of coupled S waves is sought along the common ray in the form of Equation 20, in which the amplitude factors B and C can be expressed as

$$B(\tau) = A(\tau)\mathcal{B}(\tau) \qquad C(\tau) = A(\tau)\mathcal{C}(\tau). \tag{53}$$

In Equations 53, $A(\tau)$ is again the scalar ray amplitude, see equation (34) or (36), calculated along the common S-wave ray. There are many ways how to evaluate factors $\mathcal{B}$ and $\mathcal{C}$ (Kravtsov, 1968; Pšenčík, 1998; Červený et al., 2007). Here we present a combination of coupling ray theory and of the first-order ray tracing (Farra and Pšenčík, 2010). In the approximation of Farra and Pšenčík (2010), the common ray is obtained as the first-order ray, see section on ray perturbation methods. The vectors $\mathbf{g}^{(K)}$, appearing in Equation 20, specify the first-order approximation of the S-wave polarization plane. The factors $\mathcal{B}$ and $\mathcal{C}$ in Equations 53 are then obtained as a solution of two coupled ordinary differential equations, which result from the corresponding two coupled transport equations:

$$\begin{pmatrix}\mathrm{d}\mathcal{B}/\mathrm{d}\tau \\ \mathrm{d}\mathcal{C}/\mathrm{d}\tau\end{pmatrix} = -\frac{\mathrm{i}\omega}{2}\begin{pmatrix}\mathcal{M}_{11}-1 & \mathcal{M}_{12} \\ \mathcal{M}_{12} & \mathcal{M}_{22}-1\end{pmatrix}\begin{pmatrix}\mathcal{B} \\ \mathcal{C}\end{pmatrix}. \tag{54}$$

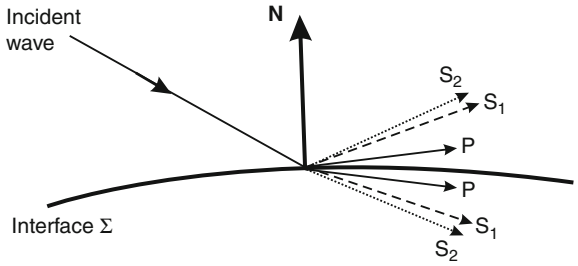
Evaluation of the matrix $\boldsymbol{\mathcal{M}}$ with elements $\mathcal{M}_{IJ}$ is simple (see Farra and Pšenčík (2010); Equations 20 and 7).

The resulting equations reduce to standard ray-theory equations in inhomogeneous isotropic media, they describe properly S-wave coupling in inhomogeneous weakly anisotropic media and even yield separate S waves when anisotropy is stronger. Common S-wave rays are regular everywhere. They do not suffer from problems well known from tracing rays of individual S waves in anisotropic media and are suitable for investigating shear-wave splitting.

Effects of structural interfaces

Assume that the ray is incident on a curved structural interface. If we wish to continue the ray computations for the reflected, transmitted, monotypic or converted waves, see Figure 5, we have to use relevant transformation relations for the ray tracing system, dynamic ray tracing system and for the ray theory amplitudes at the interface.

The transformation relations for ray tracing and dynamic ray tracing systems at interfaces are well known (Červený, 2001). For the sake of brevity, we do not present them here. We shall, however, discuss the transformation of amplitudes. In the zero-order ray approximation, the transformation of ray-theory amplitudes across an interface is described by plane-wave reflection/transmission coefficients. In other words, amplitudes of generated waves do not depend on the curvature of the wavefront of the incident wave and the curvature of the interface at the point of incidence Q. Neither do they depend on the gradients of the density and on gradients of the density-normalized elastic moduli at Q, on both sides of the interface. They depend only on the local values of the density and density-normalized elastic moduli at Q (on both sides of the interface) and on the angle of incidence (the acute angle between the slowness vector of the incident wave and the normal to the interface $\mathbf{N}$ at the point of incidence Q).



Seismic, Ray Theory, Figure 5 Slowness vectors of P, S1 and S2 waves generated at the point of incidence Q of a curved interface separating two inhomogeneous anisotropic media. All slowness vectors at Q are situated in the plane of incidence specified by the slowness vector of the incident wave and the normal to the interface **N** at Q. Ray-velocity vectors (tangent to rays) of individual waves at Q are, in general, not confined to the plane of incidence. In isotropic media, instead of reflected and transmitted S1 and S2 waves, single reflected and transmitted S waves are generated. In inhomogeneous weakly anisotropic media, single coupled S waves are generated. Ray-velocity vectors of individual waves at Q are situated in the plane of incidence in isotropic media.

Various types of R/T coefficients may be used. The displacement R/T coefficients are used most frequently (Aki and Richards, 1980; Červený et al., 1977 for isotropic media; Fedorov, 1968 for anisotropic media). Very useful are the energy R/T coefficients, as they are reciprocal. The relation between the energy R/T coefficient $\mathcal{R}(Q)$ and the displacement R/T coefficient $R(Q)$ is as follows:

$$\mathcal{R}(Q) = R(Q)\left[\frac{\rho(\tilde{Q})\mathcal{U}_n(\tilde{Q})}{\rho(Q)\mathcal{U}_n(Q)}\right]^{1/2} \tag{55}$$

(Červený 2001, sect. 5.4.3). Here Q is the point of incidence, and $\tilde{Q}$ the relevant initial point of the R/T wave, both points being, of course, identical. $\mathcal{U}_n$ is the normal component (perpendicular to the interface) of the ray-velocity vector. We further introduce the complete energy R/T coefficients $\mathcal{R}^C$ along the ray using the relation

$$\mathcal{R}^C = \prod_{k=1}^{N} \mathcal{R}(Q_k). \tag{56}$$

The complete energy R/T coefficient $\mathcal{R}^C$ corresponds to the ray which interacts N-times with interfaces (at points of incidence $Q_1, Q_2, ..., Q_N$) between the initial and end point of the ray.

Generalization of the continuation relation (36) for the ray-theory amplitudes along the ray situated in a laterally varying anisotropic medium containing curved interfaces then reads:

$$A(\tau) = \left(\frac{\rho(\tau_0)\mathcal{C}(\tau_0)|\det \mathbf{Q}^{(y)}(\tau_0)|}{\rho(\tau)\mathcal{C}(\tau)|\det \mathbf{Q}^{(y)}(\tau)|}\right)^{1/2} \mathcal{R}^C \exp[\mathrm{i}T^C(\tau,\tau_0)]A(\tau_0). \tag{57}$$

In seismic prospecting, in the technique called *amplitude variation with offset* (AVO), it is common to work with the so-called *weak-contrast R/T coefficients*. They are linearized versions of exact R/T displacement coefficients. Linearization is mostly made with respect to the contrasts of the density and elastic moduli across the interface. There is a great variety of linearized formulae depending on the type of media surrounding the interface (isotropic, anisotropic), strength of anisotropy (weak, strong), etc. The coefficients yield reasonable approximation in the vicinity of normal incidence. For increasing incidence angles, their accuracy decreases. The advantage of the weak-contrast coefficients is their simplicity and the possibility of expressing them in explicit form. The effects of the individual medium parameters on the coefficients can than be easily evaluated.

Ray-theory elastodynamic Green function

The elastodynamic Green function $G_{in}(R,t,S,t_0)$ represents the i-th Cartesian component of the displacement vector at location R and time t, due to a single-force point source situated at location S and oriented along the n-th Cartesian axis, with the time dependence $\delta(t-t_0)$. We introduce quite analogously the ray-theory elastodynamic Green function, with only two differences. The first difference is that ray-theory Green function is defined as a sum of *elementary ray-theory Green functions* computed along rays of selected elementary waves (direct, multiply reflected/transmitted, etc.). The second difference is that the elementary ray-theory Green functions are not exact, but only zero-order ray approximations.

In the frequency domain the elementary ray-theory elastodynamic Green function $G_{in}(R,S,\omega)$ for $t_0 = 0$ reads:

$$G_{in}(R,S,\omega) = \frac{g_n(S)g_i(R)\exp[\mathrm{i}T^G(R,S) + \mathrm{i}\omega T(R,S)]}{4\pi[\rho(S)\rho(R)\mathcal{C}(S)\mathcal{C}(R)]^{1/2}\mathcal{L}(R,S)}\,\mathcal{R}^C. \tag{58}$$

Here $\mathcal{L}(R,S)$ is the relative geometrical spreading, given by Equation 46, $g_i(R)$ and $g_n(S)$ are the eigenvectors of the generalized Christoffel matrix at R and S (polarization vectors corresponding to the considered elementary wave), T is the travel time along the ray from S to R, $\mathcal{R}^C$ the complete energy R/T coefficient resulting from interactions of the ray under consideration with interfaces between S and R, and $T^G(R,S)$ the complete phase shift due to caustics along the ray between S and R. The relevant KMAH index in anisotropic media may also include a contribution at a point source S (if the slowness surface of the considered wave is concave at S). In isotropic media, this contribution is always zero.

The complete energy R/T coefficient $\mathcal{R}^C$, the travel time $T(R,S)$, the relative geometrical spreading $\mathcal{L}(R,S)$ and the complete phase shift due to caustics are always reciprocal. Consequently, the elementary ray-theory

elastodynamic Green function satisfies a very important property of reciprocity:

$$G_{in}(R,S,\omega) = G_{ni}(S,R,\omega). \tag{59}$$

This relation is valid for any elementary seismic body wave generated by a point source.

For elementary ray-theory Green functions in inhomogeneous weakly anisotropic media see Pšenčík (1998).

Chaotic rays. Lyapunov exponents

In homogeneous media, geometrical spreading increases linearly with increasing length of the ray. In heterogeneous media, behavior of geometrical spreading is more complicated, and depends considerably on the degree of heterogeneity of the medium. In models, in which the heterogeneity exceeds certain degree, average geometrical spreading increases exponentially with increasing length of the ray. Rays in such a medium often exhibit chaotic behavior, which is characterized by a strong sensitivity of rays to the initial ray data (for example, to ray parameters). The rays with only slightly differing ray data at an initial point tend to diverge *exponentially* at large distances from the initial point. Consequently, the rays intersect many times and many rays pass through the same point. With such *chaotic rays*, two-point ray tracing is practically impossible, and the ray tubes are not narrow enough for travel time interpolation. The chaotic behavior of rays increases with increasing length of rays and prevents applicability of the ray theory.

The exponential divergence of chaotic rays in the phase space (the space formed by spatial coordinates x_i and slowness-vector components p_j) can be quantified by the so-called *Lyapunov exponents*. They may be introduced in several ways. It is common to express them in terms of characteristic values of the ray propagator matrix. The relevant expressions for the Lyapunov exponents and several numerical examples for 2D models without interfaces can be found in Klimeš (2002a). See also Červený et al. (2007), where other references can be found.

The estimate of the Lyapunov exponent of a single finite ray depends on its position and direction. The Lyapunov exponents associated with rays of different positions and directions can be used to calculate *average Lyapunov exponents* for the model. The average Lyapunov exponents play a very important role in smoothing the models so that they are suitable for ray tracing (Červený et al., 2007).

Ray perturbation methods

Ray perturbation methods represent an important part of the ray theory. They can be used for approximate but fast and transparent solutions of forward problems in complicated models. They play even more important role in the inverse problems.

Ray perturbation methods are useful everywhere, where we wish to compute the wavefield or its constituents (travel times, amplitudes, polarization) in complicated models, which deviate only little from simple, *reference models*, for which computations are simpler. The solutions for complicated models are then sought as perturbations of simpler solutions for the reference models. Examples are computations in weakly anisotropic media, which use an isotropic medium as reference, or in weakly dissipative media, which use a perfectly elastic medium as reference. Basic role in these approaches is played by *reference rays* traced in reference media. Solutions in perturbed media can be given in the form of a power series in the deviations of the perturbed and reference models. Mostly, the *first-order approximation*, i.e. the first term of the power series, is used.

The most frequent application of ray perturbation methods is in travel-time computations. First-order travel-time perturbation formulae for isotropic media are known and have been used (mostly in tomography) for several decades. Well known and broadly applied are also first-order travel-time formulae for anisotropic media (Červený and Jech, 1982; Hanyga, 1982; Červený, 2001, sect. 3.9). Travel-time perturbations are obtained by quadratures along reference rays. As integration parameters, the parameters along reference rays are used.

Recently, several procedures for computation of higher-order travel-time perturbations for weakly anisotropic media (note that anisotropy of the Earth is mostly weak) were proposed. The procedure based on the so-called perturbation Hamiltonians (Klimeš, 2002b; Červený et al., 2007) allows computation of highly accurate travel times along a fixed reference ray in a reference medium. Another procedure is based on the so-called first-order ray tracing described briefly below. In the latter method, second-order travel-time perturbations can be calculated along first-order rays.

Relatively recent is the use of ray perturbation methods in *first-order ray tracing* and *first-order dynamic ray tracing* (Pšenčík and Farra, 2007; Farra and Pšenčík, 2010). It allows to compute, approximately, not only rays and travel times, but whole wavefields. To derive first-order ray tracing and dynamic ray tracing, the perturbation approach is used in which deviations of anisotropy from isotropy are considered to be of the first order. Then it is just sufficient to use Equations 25 and 40 with Equation 21, in which the exact eigenvalue G_m is replaced by its first-order approximation. The resulting ray tracing provides first-order rays, first-order travel times and the first-order geometrical spreading. By simple quadratures along first-order rays, second-order travel-time corrections can be computed. This approach is applicable to P and S waves. In case of S waves, it can include the computation of coupling effects. First-order ray tracing and dynamic ray tracing are used in this case for computing common rays, first-order travel times and geometrical spreading along them, using the Hamiltonian (48). The wavefield of S waves is obtained by solving second-order coupling equations along the common rays. The procedure yields standard ray-theory results for S waves propagating in isotropic

media, and approximate results in anisotropic media when the S waves are coupled or even decoupled.

Ray perturbation method for weakly dissipative media

In viscoelastic media, the density-normalized stiffness tensor a_{ijkl} is complex valued:

$$a_{ijkl}(x_n) = a^R_{ijkl}(x_n) - \mathrm{i}a^I_{ijkl}(x_n). \tag{60}$$

If a^I_{ijkl} is small, the viscoelastic medium can be considered as a perturbation of a perfectly elastic medium (Červený, 2001, sect. 5.5.3) which has a form of the imaginary-valued term $-\mathrm{i}a^I_{ijkl}$. Reference ray in the reference perfectly elastic medium and corresponding real-valued travel time T along the reference ray between points S and R can be obtained by standard ray tracing in perfectly elastic media. The imaginary travel time T^I (travel-time perturbation due to $-\mathrm{i}a^I_{ijkl}$) can be then obtained by quadratures along the reference ray:

$$T^I = \frac{1}{2}\int_S^R Q^{-1}(\tau)\mathrm{d}\tau. \tag{61}$$

The quantity Q in Equation 61 is a direction-dependent *quality factor* for anisotropic media, corresponding to the Hamiltonian (21):

$$Q^{-1} = a^I_{ijkl}p_j p_l g_i g_k. \tag{62}$$

For general Hamiltonians, the quality factor Q is given by the relation $Q^{-1} = -\mathrm{Im}\mathcal{H}(x_i, p_j)$.

The imaginary travel time T^I in Equation 61 is responsible for the exponential amplitude decay along the reference ray. For causal dissipation, the stiffness tensor (60) is frequency dependent. The above described perturbation approach is then equivalent to the perturbation scheme, in which $a^I_{ijkl}(x_n, \omega)$ is considered to be of the order of ω^{-1} for $\omega \to \infty$ (Kravtsov and Orlov, 1990; Gajewski and Pšenčík, 1992).

In an inhomogeneous isotropic, weakly dissipative medium, the expression (62) reduces to the well-known formula

$$Q^{-1} = -\mathrm{Im}V^2/\mathrm{Re}V^2 \doteq -2\mathrm{Im}V/\mathrm{Re}V, \tag{63}$$

in which V is the complex-valued velocity, $V = \alpha$ for P waves and $V = \beta$ for S waves. Complex-valued quantities α and β are generalizations (to the complex space) of real-valued α and β from Equation 14.

Concluding remarks. Applications, modifications, and extensions of the ray method

Seismic ray method has found broad applications both in global seismology and in seismic exploration. The advantages of the seismic ray method consist in its numerical efficiency, universality, conceptual clarity, and in its ability to investigate various seismic body waves independently of other waves. Although its accuracy is only limited, the seismic ray method is the only method which is able to give an approximate answer to many problems of high-frequency seismic body wave propagation in laterally varying, isotropic or anisotropic, perfectly elastic or dissipative, layered and block structures.

In classical global seismology, the seismic ray method has been traditionally used to study the internal structure of the whole Earth, assuming that the Earth is radially symmetric. The standard Earth's model, obtained in this way, is expressed in terms of distribution of elastic velocities as a function of depth.

At present, the applications of the seismic ray method are considerably broader. It is broadly used to study the 3-D local lateral inhomogeneities in the structure, the form and physical properties of structural interfaces, the local anisotropy, attenuation, etc. In addition to forward modeling, the ray perturbation methods are also broadly used for inversions based on measured travel times or whole waveforms. In lithospheric structural studies, particularly in crustal seismology, the ray-synthetic seismograms have been also often used for ultimate comparison with observed seismograms. The computation of ray-synthetic seismograms requires determination of not only travel times, but also ray-theory amplitudes and polarization of individual waves. Seismic ray method has also found broad applications in other branches of seismology. Very important examples are the localization of seismic sources and the simultaneous localization with structural inversion.

In most applications of the ray method in seismic exploration for oil, the use of local 3-D structures with structural curved interfaces is a necessity. Sophisticated algorithms have been developed and used to image the structures under consideration. At present, the most important role is played by migration algorithms. Seismic ray theory and its extensions have found important applications in these algorithms.

The ray method is not valid universally. We have briefly described three serious limitations of the ray method: (a) The ray method can be used only for high-frequency signals. (b) In models, in which heterogeneity of the medium exceeds certain degree, the ray field has chaotic character, particularly at large distances from the source. (c) The standard ray method cannot be used for computing S waves propagating in inhomogeneous, weakly anisotropic media. It must be replaced by the coupling ray theory. The coupling ray theory must be used even in moderately or strongly anisotropic media, in the vicinity of shear-wave singular directions.

The ray method fails, however, even in other singular situations. In smooth isotropic media, the most important type of singularity are *caustics*. Caustics may attain various forms. Various extensions of the ray method can be used to compute wavefields in caustic regions. These extensions are frequency dependent. See a detailed treatment of wavefields in caustic regions in Kravtsov and Orlov (1999), and also in Stamnes (1986). In models with

smooth structural interfaces, other singularities often appear. For edge and vertex points, see Ayzenberg et al. (2007). For critical singular regions, at which head waves separate from reflected waves, see Červený and Ravindra (1971). For the waves, whose rays are tangential to interfaces, see Thomson (1989).

Specific methods, designed for different types of singularities may be used for computing wavefields in singular regions. Disadvantage of these methods is that they are different for different singularities. Morever, singular regions often overlap, and the wavefield in the overlaping region requires again different treatment. It is desirable to have available a more general extension of the ray method, applicable uniformly in any of the mentioned singular regions, or, at least, in most of them. Such an extension would simplify ray computations considerably and could even lead to more accurate results.

Several such extensions of the ray method have been proposed. We do not describe them here in detail. Instead, we merely present references, in which more details and further references can be found. Let us mention the Maslov asymptotic ray theory introduced to seismology by Chapman and Drummond (1982), see also Thomson and Chapman (1985), Chapman (2004). Another extension of the ray method is based on the summation of Gaussian beams (Popov, 1982; Červený et al., 1982). For the relation of this method with the Maslov method see Klimeš (1984). The Gaussian beam summation method has found applications both in the forward modeling of seismic wavefields and in migrations in seismic exploration. It is closely related to the method of summation of Gaussian packets (Červený et al., 2007). Ray theory can be also used in the Born scattering theory (Chapman and Coates, 1994; Chapman, 2004). For waves reflected from a smooth structural interface separating two heterogeneous, isotropic or anisotropic media, the Kirchhoff surface integral method can be used. For details and many references see Chapman (2004, sect. 10.4). Another useful extension of the ray method is the one-way wave equation approach (Thomson, 1999).

Acknowledgments

The authors are very grateful to Luděk Klimeš and Ravi Kumar for valuable comments and recommendations. The research was supported by the consortium project Seismic Waves in Complex 3-D Structures, by research projects 205/07/0032 and 205/08/0332 of the Grant Agency of the Czech Republic; and by research project MSM0021620860 of the Ministry of Education of the Czech Republic.

Bibliography

Aki, K., and Richards, P., 1980. *Quantitative Seismology*. San Francisco: WH Freeman.

Ayzenberg, M. A., Aizenberg, A. M., Helle, H. B., Klem-Musatov, K. D., Pajchel, J., and Ursin, B., 2007. 3D diffraction modeling of singly scattered acoustic wavefields based on the combination of surface integral propagators and transmission operators. *Geophysics*, **72**, SM19–SM34.

Babich, V. M., 1956. Ray method of the computation of the intensity of wave fronts (in Russian). *Doklady Akademii Nauk SSSR*, **110**, 355–357.

Babich, V. M., 1961. Ray method of the computation of the intensity of wave fronts in elastic inhomogeneous anisotropic medium. In Petrashen, G. I. (ed.), *Problems of the Dynamic Theory of Propagation of Seismic Waves* 77 (in Russian). Leningrad: Leningrad University Press, Vol. 5, pp. 36–46. Translation to English: Geophysical Journal International., 118: 379–383, 1994.

Bakker, P. M., 1998. Phase shift at caustics along rays in anisotropic media. *Geophysical Journal International*, **134**, 515–518.

Bakker, P. M., 2002. Coupled anisotropic shear-wave ray tracing in situations where associated slowness sheets are almost tangent. *Pure and Applied Geophysics*, **159**, 1403–1417.

Bulant, P., and Klimeš, L., 2002. Numerical algorithm of the coupling ray theory in weakly anisotropic media. *Pure and Applied Geophysics*, **159**, 1419–1435.

Červený, V., 1972. Seismic rays and ray intensities in inhomogeneous anisotropic media. *Geophysical Journal of Royal Astronomical Society*, **29**, 1–13.

Červený, V., 2001. *Seismic Ray Theory*. Cambridge: Cambridge University Press.

Červený, V., and Jech, J., 1982. Linearized solutions of kinematic problems of seismic body waves in inhomogeneous slightly anisotropic media. *Journal of Geophysics*, **51**, 96–104.

Červený, V., and Ravindra, R., 1971. *Theory of Seismic Head Waves*. Toronto: Toronto University Press.

Červený, V., Molotkov, I. A., and Pšenčík, I., 1977. *Ray Method in Seismology*. Praha: Univerzita Karlova.

Červený, V., Popov, M. M., and Pšenčík, I., 1982. Computation of wave fields in inhomogeneous media. Gaussian beam approach. *Geophysical Journal of Royal Astronomical Society*, **70**, 109–128.

Červený, V., Klimeš, L., and Pšenčík, I., 1988. Complete seismic ray tracing in three-dimensional structures. In Doornbos, D. J. (ed.), *Seismological Algorithms*. New York: Academic, pp. 89–168.

Červený, V., Klimeš, L., and Pšenčík, I., 2007. Seismic ray method: recent developments. *Advances in Geophysics*, **48**, 1–126. http://www.sciencedirect.com/science/bookseries/00652687.

Chapman, C. H., 2002. Seismic ray theory and finite frequency extensions. In Lee, W. H. K., Kanamori, H., and Jennings, P. C. (eds.), *International Handbook of Earthquake and Engineering Seismology, Part A*. New York: Academic, pp. 103–123.

Chapman, C. H., 2004. *Fundamentals of Seismic Wave Propagation*. Cambridge: Cambridge University Press.

Chapman, C. H., and Coates, R. T., 1994. Generalized Born scattering in anisotropic media. *Wave Motion*, **19**, 309–341.

Chapman, C. H., and Drummond, R., 1982. Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory. *Bulletin of the Seismological Society of America*, **72**, S277–S317.

Coates, R. T., and Chapman, C. H., 1990. Quasi–shear wave coupling in weakly anisotropic 3-D media. *Geophysical Journal International*, **103**, 301–320.

Farra, V., and Pšenčík, I., 2010. Coupled S waves in inhomogeneous weakly anisotropic media using first-order ray tracing. *Geophysical Journal International*, **180**, 405–417.

Fedorov, F. I., 1968. *Theory of Elastic Waves in Crystals*. New York: Plenum.

Gajewski, D., and Pšenčík, I., 1992. Vector wavefield for weakly attenuating anisotropic media by the ray method. *Geophysics*, **57**, 27–38.
Gjøystdal, H., Iversen, E., Laurain, R., Lecomte, I., Vinje, V., and Åstebol, K., 2002. Review of ray theory applications in modelling and imaging of seismic data. *Studia Geophsica et Geodaetica*, **46**, 113–164.
Hanyga, A., 1982. The kinematic inverse problem for weakly laterally inhomogeneous anisotropic media. *Tectonophysics*, **90**, 253–262.
Jech, J., and Pšenčík, I., 1989. First-order perturbation method for anisotropic media. *Geophysical Journal International*, **99**, 369–376.
Karal, F. C., and Keller, J. B., 1959. Elastic wave propagation in homogeneous and inhomogeneous media. *The Journal of the Acoustical Society of America*, **31**, 694–705.
Klimeš, L., 1984. The relation between Gaussian beams and Maslov asymptotic theory. *Studia Geophysica et Geodaetica*, **28**, 237–247.
Klimeš, L., 2002a. Lyapunov exponents for 2-D ray tracing without interfaces. *Pure and Applied Geophysics*, **159**, 1465–1485.
Klimeš, L., 2002b. Second-order and higher-order perturbations of travel time in isotropic and anisotropic media. *Studia Geophysica et Geodaetica*, **46**, 213–248.
Klimeš, L., 2006. Common-ray tracing and dynamic ray tracing for S waves in a smooth elastic anisotropic medium. *Studia Geophysica et Geodaetica*, **50**, 449–461.
Klimeš, L., 2010. Phase shift of the Green tensor due to caustics in anisotropic media. *Studia Geophysica et Geodaetica*, **54**, 269–289.
Kravtsov, Yu. A., 1968. "Quasiisotropic" approximation to geometrical optics. *Doklady Akademii Nauk SSSR*, **183**(1), 74–77 (in Russian).
Kravtsov, Yu. A., and Orlov, Yu. I., 1990. *Geometrical Optics of Inhomogeneous Media*. Heidelberg: Springer Verlag.
Kravtsov, Yu. A., and Orlov, Yu. I., 1999. *Caustics, Catastrophes and Wave Fields*. Heidelberg: Springer Verlag.
Popov, M. M., 1982. A new method of computation of wave fields using Gaussian beams. *Wave Motion*, **4**, 85–97.
Pšenčík, I., 1998. Green's functions for inhomogeneous weakly anisotropic media. *Geophysical Journal International*, **135**, 279–288.
Pšenčík, I., and Farra, V., 2007. First-order P-wave ray synthetic seismograms in inhomogeneous weakly anisotropic media. *Geophysical Journal International*, **170**, 1243–1252.
Stamnes, J. J., 1986. *Waves in Focal Regions*. Bristol: Adam Hilger.
Thomson, C. J., 1989. Corrections for grazing rays to 2-D seismic modelling. *Geophysical Journal International*, **96**, 415–446.
Thomson, C. J., 1999. The gap between seismic ray theory and full wavefield extrapolation. *Geophysical Journal International*, **137**, 364–380.
Thomson, C. J., and Chapman, C. H., 1985. An introduction to Maslov's asymptotic method. *Geophysical Journal of Royal Astronomical Society*, **83**, 143–168.
Virieux, J., 1996. Seismic ray tracing. In Boschi, E., Ekström, G., and Morelli, A. (eds.), *Seismic Modelling of Earth Structures*. Bologna: Editrice Compositori, pp. 223–304.

Cross-references

Energy Partitioning of Seismic Waves
Seismic Anisotropy
Seismic Diffraction
Seismic Waves, Scattering
Seismic, Migration
Seismic, Viscoelastic Attenuation
Seismic, Waveform Modeling and Tomography
Traveltime Tomography Using Controlled-Source Seismic Data

SEISMIC, RECEIVER FUNCTION TECHNIQUE

Rainer Kind[1], Xiaohui Yuan[2]
[1]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 2.4, Seismology, Potsdam, Germany
[2]Section Seismology, GFZ German Research Centre for Geosciences, Potsdam, Germany

Synonyms

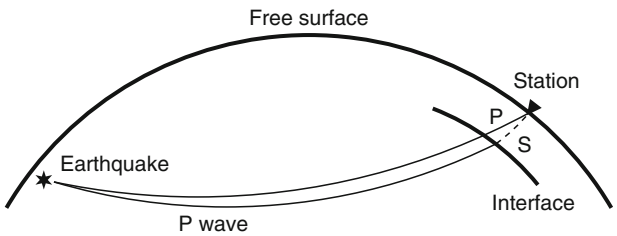
Receiver functions; Scattered teleseismic waves

Definition

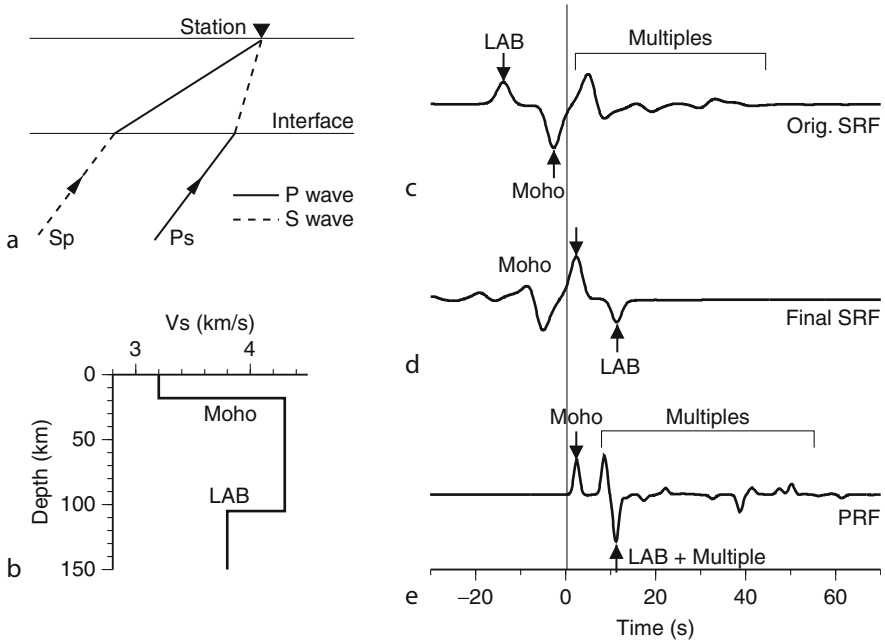
Receiver function. Response of the Earth's structure below seismic stations to incident teleseismic waves. Seismic discontinuities in the Earth's crust or upper mantle cause waves from distant earthquakes to be converted from P to S waves or vice versa, or to be multiply reflected between discontinuities before arriving the stations. Such scattered waves carry information about the seismic structure of the lithosphere and upper mantle. Scattered waves are weak signals and summation of many records is usually required to visualize the useful information.

Receiver function technique

If a seismic wave hits a discontinuity between two solid materials with different physical parameters, a part of the wave will be reflected and another part will be transmitted. There is also mode conversion between compressional (P) waves and shear (S) waves (see *Energy Partitioning of Seismic Waves*). Therefore, a P wave crossing a discontinuity will generate an S wave (called Ps), which will follow the P wave with a slower speed; and an S wave will produce a P wave (called Sp), which will run ahead of the S wave with a faster speed. If these wave types are recorded at the surface and their incidence angles and the material velocities are known, the depth of the discontinuity, where they are generated can be determined from the differential times of the two seismic phases, mother phase



Seismic, Receiver Function Technique, Figure 1 P receiver function geometry. The target zone below the three-component receiver is illuminated from below. The source is an earthquake at a large epicentral distance. When a plane P wave encounters interfaces below the receiver, Ps waves are generated. The incidence angle of the P wave is always bigger than that of the Ps waves. The travel path of S receiver functions (SRFs) is very similar, except that the incidence angle of the S wave is smaller than that of the Sp wave.

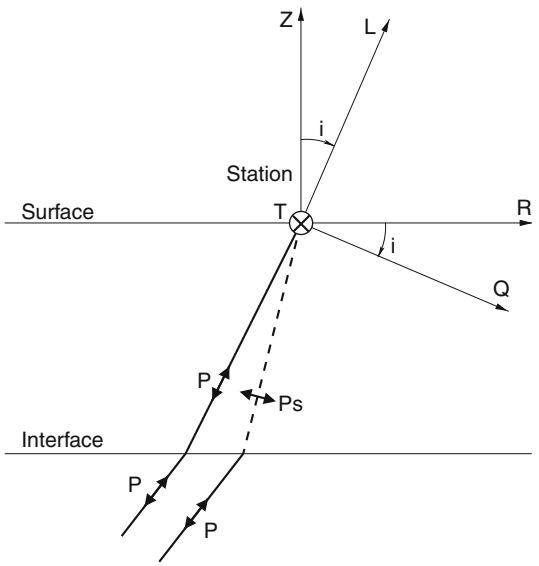


Seismic, Receiver Function Technique, Figure 2 (**a**) Sketch showing ray paths of teleseismic Ps and Sp converted waves below a station. Note that the piercing point of Sp is offset farther away from the station than that of Ps. (**b**) Velocity model for the calculation of P and S receiver functions. (**c**) Theoretical S receiver function (SRF), primary conversions arrive prior to S (which arrives at zero time and is omitted here), crustal multiples arrive after S. (**d**) S receiver function (trace **c**) after reversal of timescale and amplitude for better comparison with P receiver functions (PRFs). (**e**) P receiver functions (PRFs) of model in (**b**). The main difference of P and S receiver functions is that S receiver functions are free of crustal multiples at the time of the lithosphere–asthenosphere boundary (LAB) arrivals. Note that in the case of a homogeneous halfspace, no signal would be contained in the above traces (Yuan et al., 2007).

and daughter phase. The waveforms of both phases carry information about the structure of the discontinuity, which can be isolated by the receiver function technique. The ray path of such waves is shown in Figure 1. Figure 2 compares incident P and S phases for a simple model with two discontinuities (Moho and lithosphere–asthenosphere boundary, LAB, see *Earth's Structure, Continental Crust*; *Lithosphere, Continental*; *Lithosphere, Oceanic*), which cause converted and multiply reflected phases that lead finally to the receiver functions. There are a number of processing steps required in the receiver function technique, which will be considered in the following.

Separation of P and S waves

Since the converted daughter phase is a different wave type with different particle motion, it can be separated from the mother phase. The usual vertical, north–south and east–west (ZNE) components of a seismic record should be rotated, for better identification of signals, into a coordinate system where P, SV, and SH wave types are on different components. Due to heterogeneities close to the station, such a coordinate system (frequently called LQT or P-SV-SH system, see Figure 3) depends on the local angle of incidence and back azimuth. In many cases, only the horizontal components N and E are rotated into the radial (R) and transverse (T) components. Frequently

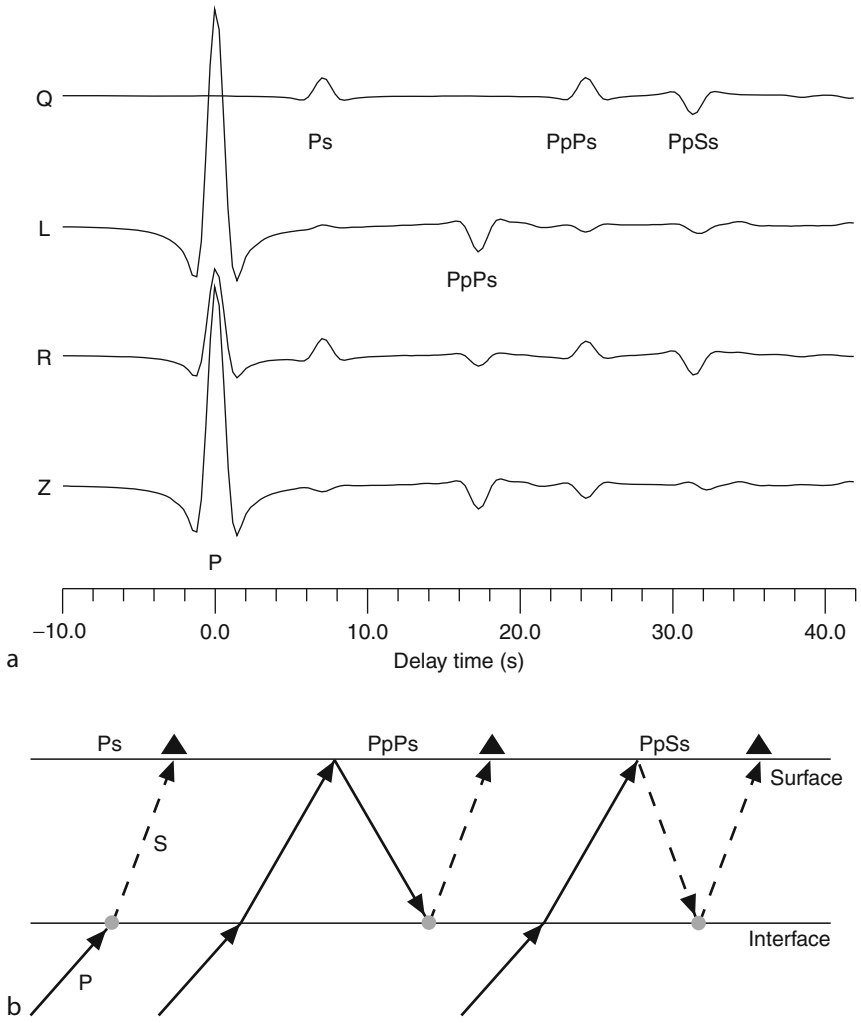


Seismic, Receiver Function Technique, Figure 3 Rotation of the radial (R) and vertical (Z) components around the angle of incidence into the L and Q system (Vinnik, 1977). The P wave is in this case only on the L component and the SV wave is only on the Q component. The SH wave remains on the T component.

the rotation is done according to the theoretical back azimuth between receiver and source and the theoretical angle of incidence according to epicentral distance. A local ray coordinate system may be determined from the diagonalization of the covariance matrix of the ZNE recordings within a certain time window (Montalbetti and Kanasewich, 1970; Kind and Vinnik, 1988). The length of the time window used plays an important role for determination of rotation angles with this technique. The influence of the free surface of the Earth is small and in most cases not considered. In Figure 4, theoretical seismograms in the ZR and LQ coordinate system are shown, along with ray diagrams for direct Ps phase and multiple phases PpPs and PpSs.

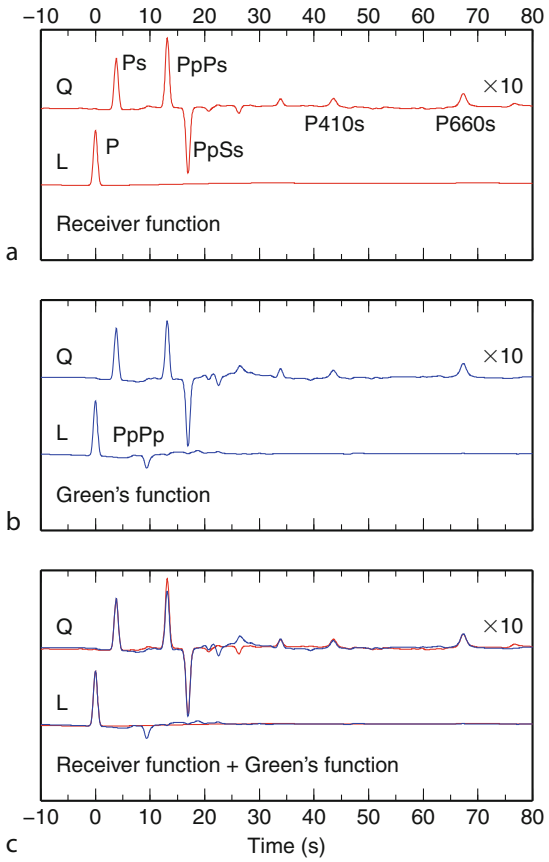
Deconvolution

After separation of P and SV wave types on separate components, it is much easier to identify and interpret the small converted signals. In earlier times when only analog data records were available, only a few records could be searched for converted phases. With the large amount of digital data now available (see *Seismological Networks*), weak signals can be enlarged relatively easily by summation of many records. A number of problems need to be considered before seismic records can be summed. Different waveforms and amplitudes generated by different earthquakes are another problem, which needs to be taken into account before summation of many traces can be applied. After rotation, records of one source and many distributed stations may be summed because all records are caused by a single event with the same source-time function. This is the well-known delay and sum technique. However the results of such a summation will be the response function at the source region (including, e.g., depth phases) and not the "receiver function." If we are interested in the structure beneath a seismic receiver,



Seismic, Receiver Function Technique, Figure 4 (**a**) Effects of rotation of theoretical Z and R coordinates into the LQ system. Phases arriving as S wave (Ps, PpPs, PpSs) at the station remain on the Q component only. Phases that arrive as P waves remain on the L component (P, PpPp). (**b**) Ray paths of crustal multiples.

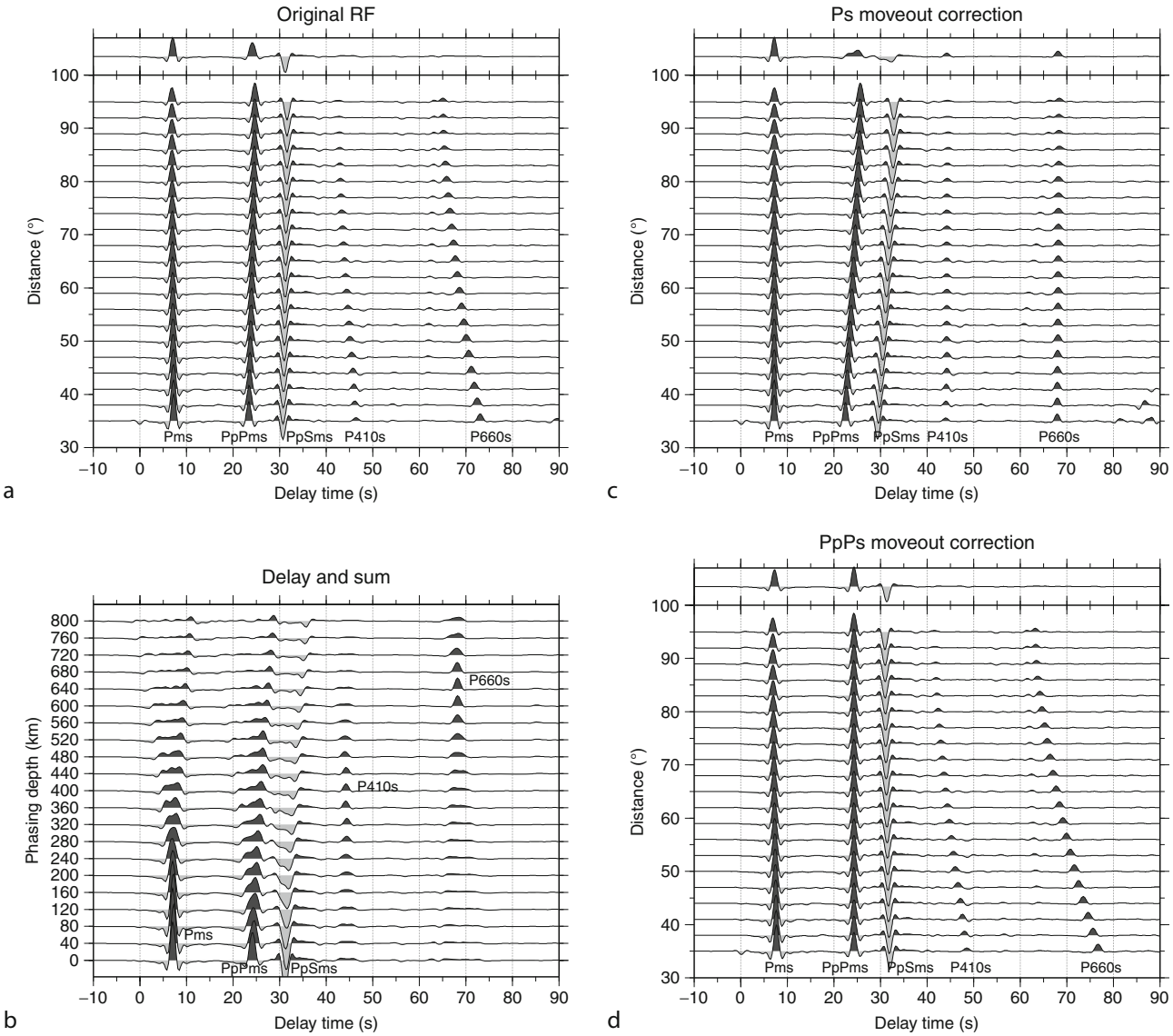
records from many earthquakes at this receiver with very different source-time functions and magnitudes must be summed. Only in this case we will obtain as the result of the summation the response of the structure beneath a receiver. Influences of the receiver structure are common to all records and will be enhanced by summation. In contrast, influences of the source or the source structure (depth phases, conversions near the source) will be reduced. Obtaining the structure beneath a receiver is especially interesting because in contrast to sources we can move the receivers where we need them and record teleseismic events to study the structure at nearly any location (with many problems at ocean bottom stations). The procedure used for source equalization is usually deconvolution. In the frequency domain, deconvolution of P receiver functions (PRFs) means division of the SV component by the P component and in S receiver functions (SRFs) it means division of the P component by the SV component (Langston, 1979; Gurrola et al., 1995). In time domain, a Wiener filtering approach (Kind et al., 1995) can be applied. For P receiver functions, an inverse filter can be generated by minimizing the least-square difference between the observed P waveform on the vertical component and the desired delta-like spike function. The inverse filter is then convolved with other components to obtain receiver functions. Ligorria and Ammon (1999) proposed an alternative approach to iteratively remove the source and propagation effects. The question which type of deconvolution is carried out does not seem to be very significant. After deconvolution, amplitude normalization is applied, resulting in amplitude values of the converted phase in percent of the incident phase. Deconvolution, however, has one principal disadvantage. One entire component is considered as the source-time function. For example, the entire P component over duration of perhaps 100 s is deconvolved from the SV component. This means all signals within this window on the P component are considered source. For example, P multiples within the crust at the receiver site will also be eliminated by this kind of deconvolution. In this sense, deconvolution prevents computation of the complete Green's function (impulse response) at the receiver site (Kumar et al., 2010; Baig et al., 2005; Bostock, 2004; Langston and Hammer, 2001). The modification of seismic records caused by deconvolution is demonstrated in Figure 5. Kumar et al. (2010) have shown that plain summation of many records of a seismic station results in SV component traces, which are nearly identical with deconvolved traces and P component traces are preserved. They only applied amplitude and sign equalization and summed all traces aligned along the maximum of the P signal (see Kumar et al., 2010). The only disadvantage of the new technique is that more traces are needed to obtain the same signal-to-noise ratio.



Seismic, Receiver Function Technique, Figure 5
(**a**) Deconvolved, (**b**) original theoretical seismograms (Kind, 1985) and (**c**) the overlay of both for a simple crust-mantle model. All phases except P are eliminated on the deconvolved P component. The deconvolved SV (Q) component differs significantly from the theoretical SV component. The P_MP phase has disappeared on the deconvolved SV component, as expected. However, the PpPs Moho multiple reflection has a significantly distorted amplitude (Kumar et al., 2010).

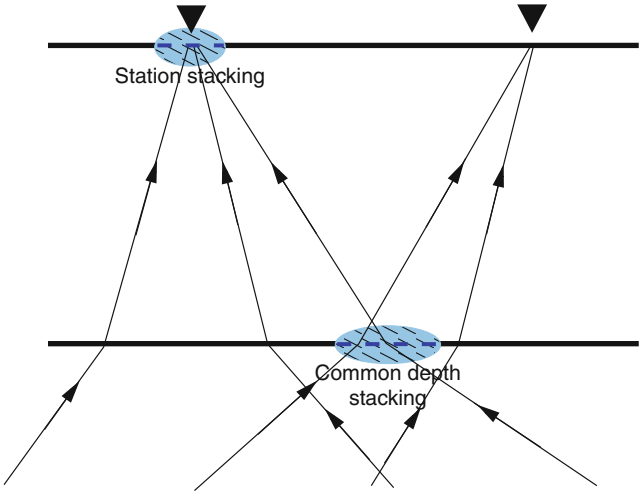
Moveout correction and summation

An important problem is that seismic phases travel with different velocities. This means that the differential times of seismic phases in one record depend on epicentral distances. Therefore, only records from similar epicentral distances can be summed. Traces from different distances may also be summed if the conversions arrive very closely in time to their mother phase (perhaps still possible for crustal conversions). For converted phases arriving later from greater depths, it is necessary to consider effects of their different slowness. Vinnik (1977) has solved this problem by using a delay and sum technique known from array processing and applying it to the upper mantle discontinuities. Results are displayed in a slowness-time plot (or conversion-depth-time plot, see Figure 6). Another solution is that the distance moveout correction procedure could be applied before summation, which is very well known in applied seismics (see *Deep Seismic Reflection and Refraction Profiling*). In moveout correction, the timescale is stretched or compressed in order to parallelize the same type of traveltime curves. A fixed reference



Seismic, Receiver Function Technique, Figure 6 Theoretical P receiver functions in original form and moveout corrected. A velocity model is needed to apply the moveout correction. (*Bottom left*) Delay and sum technique developed by Vinnik (1977). Delays are computed for a number of assumed depths of conversion.

slowness of 6.4 s/degree (or 67° epicentral distance) is frequently used at which the timescale remains unchanged. At smaller distances the timescale is compressed, and expanded at larger distances, in order to transform the traveltime curve of the converted phase under investigation into one parallel to the mother signal. After moveout correction, a number of traces may be lined up and summed into a single trace. A group of traces used for summation could be all records from either one station or records with piercing points in a certain geographical region (box) at a certain depth (common conversion point technique, Dueker and Sheehan, 1997, see Figure 7). The advantage of the moveout correction and summation technique is that single traces representing groups of traces can be easily compared. After moveout correction for direct conversions, the summation traces contain only such phases. Surface multiples are destructively superimposed because of their different slowness. Moveout corrections cannot only be applied to Ps or Sp conversions, but also to other phases like multiples. In this case, the summation trace contains only a certain type of multiples (see Figure 6). Moveout correction and summation also solves a special problem in S receiver functions. Not every precursor of S on the P component is an S-to-P



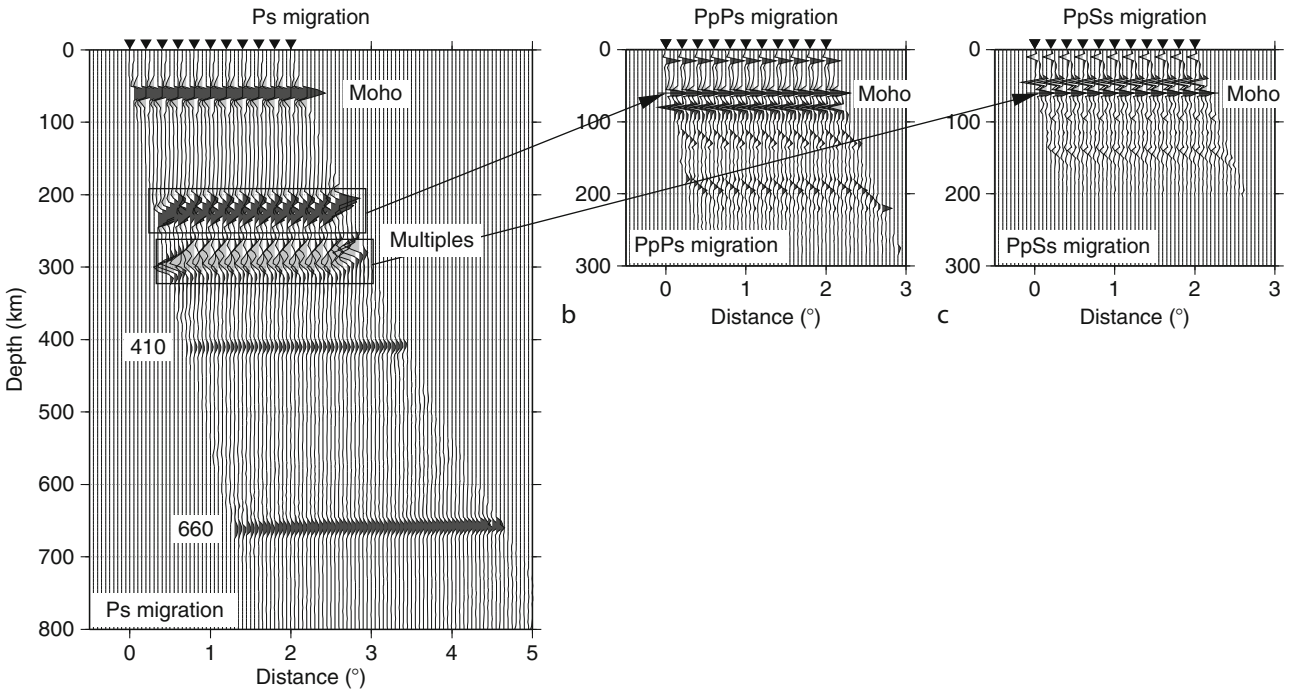
Seismic, Receiver Function Technique, Figure 7 Cartoon showing receiver function rays in a simple Earth model. Single station summation is useful for shallower studies or for averages over larger regions in greater depths. In case of dense deployments of several stations, common depth summation, (CCP) may be used.

conversion. Surface multiples of the P phase can also arrive on the vertical component as precursors of S (Bock, 1994). These phases have, fortunately, a different slowness and are suppressed in the delay and sum technique or the moveout correction and sum technique.

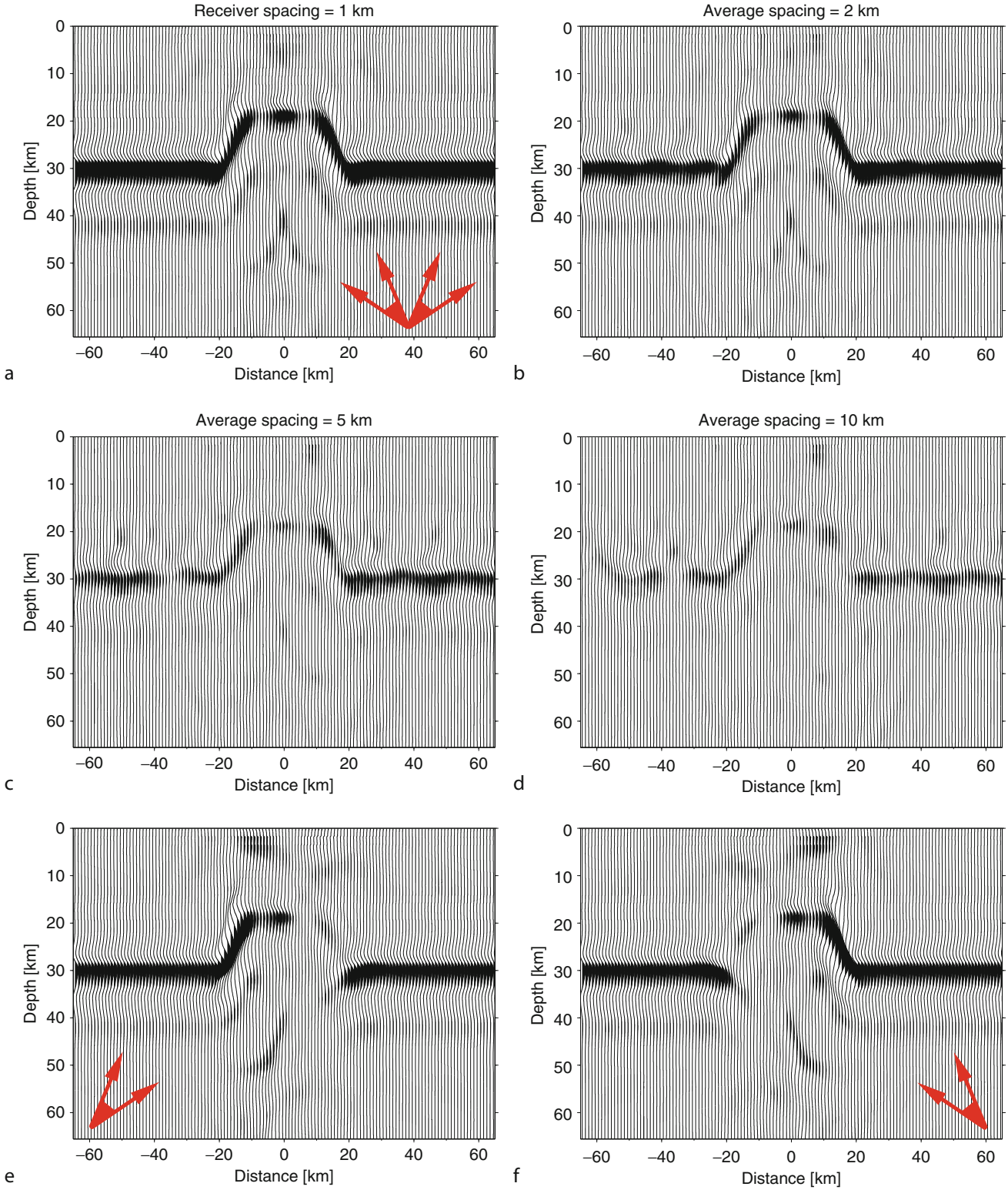
CCP stack and migration

Stacking by single stations is meaningful if piercing points from different stations are not overlapping in the depth range studied. If several stations are closely spaced and piercing points are overlapping at the study depth, summation by common regions of piercing points at that depth may be more meaningful (common conversion point stacking, CCP). In this case, piercing point locations for conversions at a defined depth are computed and all traces with piercing points inside a defined area are summed (see Figure 7).

Stacked time domain receiver functions are not yet the goal of the processing, even if clear converted phases can be recognized. We need to know where in the three-dimensional Earth the conversion occurs. Therefore, the final goal is the migration from the time series into



Seismic, Receiver Function Technique, Figure 8 Principle of depth migration of seismic records (theoretical seismograms in this case). Each ray from each station is back projected into an assumed model of the Earth. The amplitudes of the seismogram are distributed along this ray path. In the left figure it is assumed that all energy in the seismogram originates from direct conversions; in the other figures it is assumed that all energy is caused by multiples. Migration can, like moveout correction, only be done for one wave type at a time. The energy of the other phase type is migrated to the wrong depth and not focused.



Seismic, Receiver Function Technique, Figure 9 (**a**–**d**) Migrated theoretical 1-Hz receiver functions of a dome structure with different receiver spacing. A spacing of 5 or less kilometers is required to obtain a good image of the structure. (**e**–**f**) Effects of illuminating the structure from limited back azimuths (see *arrows*, Ryberg and Weber, 2000). Station spacing is 1 km.

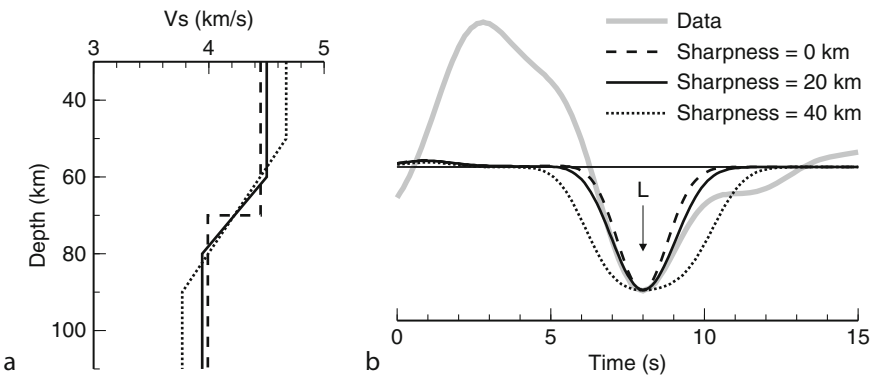
a depth model (see *Seismic, Migration*). Depth migration with a known velocity depth model is necessary for this purpose. Migration is a very well-known technique in controlled source seismics. In earthquake seismology with relatively sparse station distribution more simple versions of migration with one-dimensional models are often used. The amplitudes of the receiver function traces are distributed along the ray path of a known velocity depth model. The Earth model is divided in boxes and amplitudes of all rays in one box are summed. The result is an approximate two-dimensional (or three-dimensional) distribution of seismic energy in space where hopefully seismic discontinuities can be identified (see Figure 8). In the case of P receiver functions, such images are blurred with multiples from the surface, which produce apparent discontinuities. One of the first examples of receiver function depth migration is given by Kosarev et al. (1999). A more sophisticated development is given, e.g., by Bostock (2002).

The results of the processing steps described so far indicate the great similarity with steep angle–controlled source techniques. The topography of seismic discontinuities in the interior of the Earth is displayed either in the time or space domain. Distance moveout corrections and depth migration are the most important common steps. The main differences are the frequency content (which determines the resolution) and depth penetration. Steep angle seismics uses frequencies of many Hertz, whereas the teleseismic signals have periods from one to many seconds. Steep angle seismics does not always reach the Moho, whereas receiver functions have practically unlimited depth penetration since the signals are incident from below. Ryberg and Weber (2000) conclude from computations of theoretical seismograms of two-dimensional models that a station spacing of a few kilometers is required if structures with a few kilometers scale length need to be resolved (see Figure 9). A station density of a few kilometers has so far been achieved only in very few deployments (e.g., Hansen and Dueker, 2009).

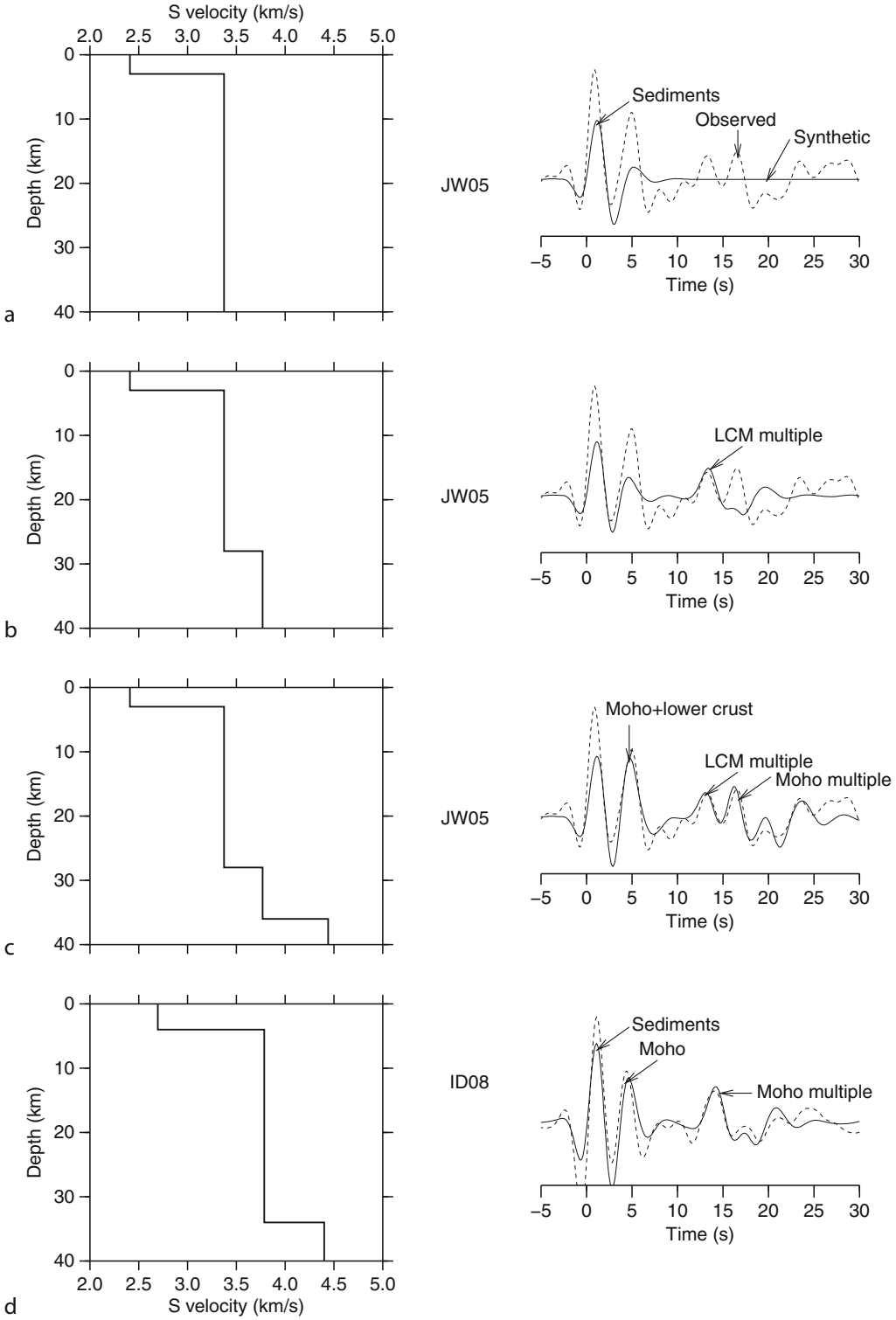
Waveform modeling

The above described processing steps for receiver functions lead to the detection and location of discontinuities in the interior of the Earth and their topography, ranging from the base of the sedimentary layers to the upper mantle transition zone between 410- and 660-km depth. A frequently applied additional step in receiver functions is the inversion of the waveforms over the entire length of the receiver function at a station into a one-dimensional velocity depth model (e.g., Ammon et al., 1990; Kind et al., 1995). Such an inversion technique has the capability of fitting the data nearly perfectly. However, problems involved are the nonuniqueness of the inversion and the assumption that seismic phases, which have been generated by a three-dimensional Earth, can be transformed into a one-dimensional model (Ammon et al., 1990). Although the nonuniqueness problem has been improved recently by using long-period signals (Jacobsen and Svenningsen, 2008), even a perfect fit of the observed waveforms with theoretical waveforms encounters the problem of fitting three-dimensional effects and possible noise into a one-dimensional model.

Somewhat closer to reality seems to be the inversion of isolated waveforms of conversions from individual discontinuities in P or S receiver functions (e.g., Moho or LAB). An example is given by Li et al. (2007), see Figure10. They summed about 8,000 S-to-P converted signals from the LAB (S receiver functions) in the western US and inverted the width of the summed signal in terms of thickness of the LAB gradient. They obtained a transition zone of not more than 20 km. However, determinations of the size of the velocity jump and of the thickness of the LAB from summation traces may depend on



Seismic, Receiver Function Technique, Figure 10 Modeling of the LAB summation phase in the western US (Li et al., 2007, Figure 8). (**a**) Models with different sharpness and velocity contrast of the LAB. Dashed line: LAB as a first-order discontinuity; solid line: LAB as a 20-km thick gradient transition zone and dotted line: LAB as a 40-km thick gradient transition zone. (**b**) Waveforms of data and synthetic S receiver functions. The thick gray line is the summation of all the S receiver function stacks for each station after an alignment of the LAB phase. A 3–20 s bandpass filter is applied prior to alignment and summation for a better phase correlation. Synthetic S receiver functions correspond to the models shown in (**a**). The large positive swing in the data at less than 5 s is the averaged Moho signal, which is not computed in the synthetic seismograms.



Seismic, Receiver Function Technique, Figure 11 Forward waveform modeling of P receiver functions with relatively simple models (Mohsen et al., 2005).

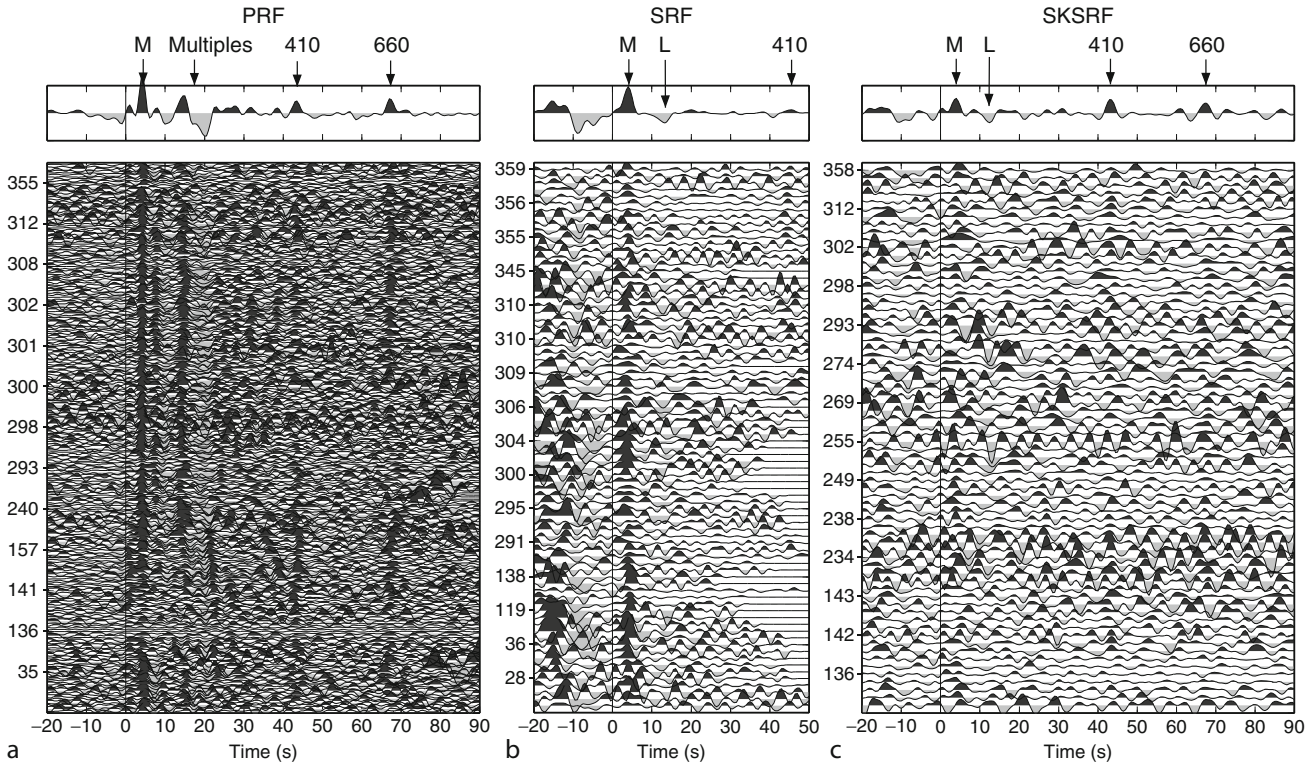
focusing and defocusing effects of the topography within the region of piercing points and erroneous transform topography into the vertical velocity structure. Therefore, this 20-km gradient must be considered as an upper limit in this case. Another example is given by Mohsen et al. (2005) in Figure 11. They show how each one of the observed seismic phases may be modeled with good accuracy with a relatively simple model.

Two data examples

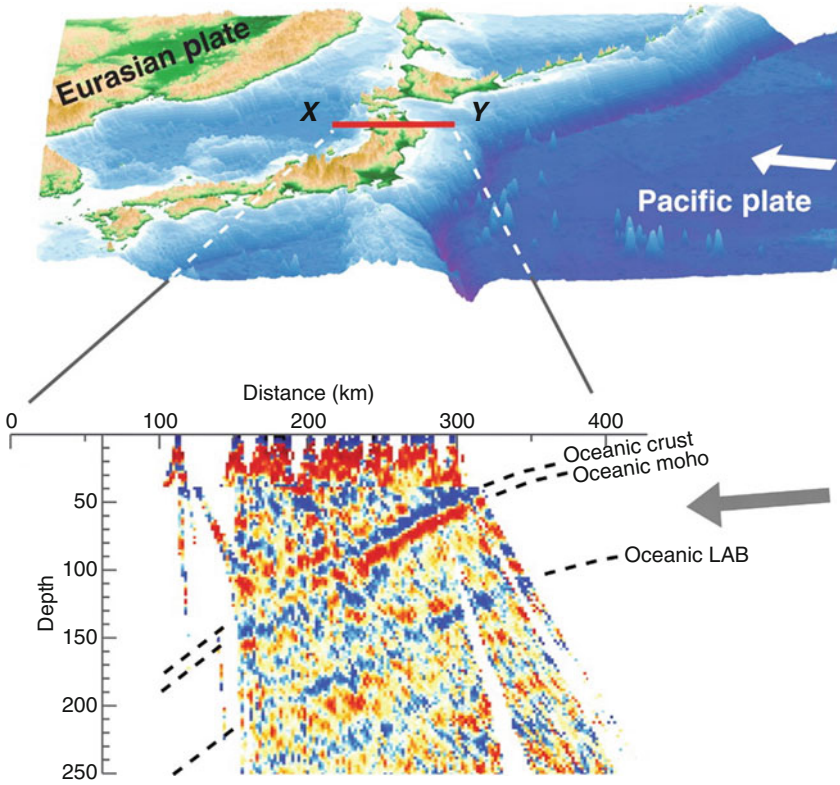
In Figure 12 are shown P, S, and SKS receiver functions in time domain from a seismically quiet station in northern Canada. The interesting point in these data is that in S and SKS receiver function (SKSRF), a clear negative phase is observed at 12–13 s, which corresponds to a depth of about 110 km. A negative signal indicates a low velocity zone. Such a zone in that depth may be caused by the LAB. However, from surface wave studies a much larger depth of the LAB is expected in this region. These contradicting results pose very interesting questions about the structure of the upper mantle beneath old cratons. In Figure 13 is shown an excellent example of depth migrated P receiver functions beneath a subduction zone (Kawakatsu et al., 2009). The subducting oceanic lithosphere is completely visible, including upper and lower boundary of the oceanic crust and the lower boundary of the oceanic lithosphere (see *Lithosphere, Oceanic*).

Summary

Receiver functions or converted waves belong to a class of small secondary scattered waves generated by a relatively strong main phase like P or S. Similar scattered phases are precursors of other main phases like PP, SS, or P′P′, which have the great advantage to cover areas that are difficult to access, like oceans (see *Body Waves*; *Seismic Phase Names: IASPEI Standard*). These techniques are since about 30 years the main tools to study discontinuities of material properties in the Earth. The other main seismic tool is tomography, which is sensitive to smooth changes of material properties. The success of the scattered wave techniques was made possible by the great extension of high quality seismic networks in the last decades. Permanent and mobile networks supplement each other. Especially studies of the lithospheric plates and the mantle transition zone have gained much from the new techniques. Many more detailed results can be expected when much denser networks will be used in future.



Seismic, Receiver Function Technique, Figure 12 PRFs, SRFs, and SKSRFs of station YKW3 (Yellow Knife, Canada). The receiver functions are corrected for moveout with a reference slowness of 6.4 s/degree and a reference model of the IASP91 (Kennett and Engdahl, 1991). The number of the PRFs is more than 200, the numbers of the SRFs and SKSRFs are ~80 each. A three-pole low-pass filter with corner frequency of 3 S is applied to all the data. Back azimuth values for some sample traces are labeled to the left of each trace. The summation trace is plotted above each receiver function section. Main conversion phases are labeled on the summation traces.



Seismic, Receiver Function Technique, Figure 13 P receiver functions for ocean bottom station WP2 and receiver function image of subducting Pacific plate beneath northeast Japan. In the topographic map, the inverted red triangle indicates the location of WP2, which is deployed at a depth of 460 m below the seafloor with a water column of 5,566-m thick. (**a**) P receiver function image along the profile XY using dense land seismic data of Hi-net from Japan. *Red* and *blue* colors indicate velocity increase (from shallow to deep) and decrease at the point, respectively. The top surface of the slab and the oceanic Moho are clearly imaged, as well as the bottom surface of the slab (i.e., subducting LAB). (**b**) P receiver function and S receiver function for WP2. Negative phases associated with a shear wave velocity drop marked as LAB appear to correspond to the strong negative (*blue*) signature of the subducting slab below Japan. (After Kawakatsu et al., 2009.)

A recent review of the receiver function technique has been given by Rondenay (2009). A number of additional processing steps are the use of multiples to determine average crustal properties (e.g., Zhu and Kanamori, 2000), the role of anisotropy (e.g., Levin and Park, 1997; Schulte-Pelkum et al., 2005), or three-dimensional inversion (e.g., Bostock, 2002).

Bibliography

Ammon, C. J., Randall, G. E., and Zandt, G., 1990. On the nonuniqueness of receiver function Inversions. *Journal of Geophysical Research*, **95**, 15303–15318.

Baig, A. M., Bostock, M. G., and Mercier, J.-P., 2005. Spectral reconstruction of teleseismic Green's functions. *Journal of Geophysical Research*, **110**, B08306.

Bock, G., 1994. Multiples as precursors of S, SKS and ScS. *Geophysical Journal International*, **119**(2), 421–427.

Bostock, M. G., 2002. Kirchhoff-approximate inversion of teleseismic wave fields. *Geophysical Journal International*, **149**, 787–795.

Bostock, M. G., 2004. Green's functions, source signatures, and the normalization of teleseismic wave fields. *Journal of Geophysical Research*, 109.

Dueker, K. G., and Sheehan, A. F., 1997. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hotspot track. *Journal of Geophysical Research*, **102**, 8313–8327.

Gurrola, H., Baker, G. E., and Minster, J. B., 1995. Simultaneous time-domain deconvolution with application to the computation of receiver functions. *Geophysical Journal International*, **120**, 537–543.

Hansen, S., and Dueker, K., 2009. P- and S-Wave Receiver Function Images of Crustal Imbrication beneath the Cheyenne Belt in Southeast Wyoming. *Bulletin. Seismological Society of America*, **99**, 1953–1961, doi:10.1785/0120080168.

Jacobsen, B. H., and Svenningsen, L., 2008. Enhanced uniqueness and linearity of receiver function inversion. *Bulletin. Seismological Society of America*, **98**, 1756–1767.

Kawakatsu, H., Kumar, P., Takei, Y., Shinohara, M., Kanazawa, T., Araki, E., and Suyehiro, K., 2009. Seismic evidence for sharp lithosphere-asthenosphere boundaries of oceanic plates. *Science*, **324**, 499–502.

Kennett, B. L. N., and Engdahl, E. R., 1991. Travel times for global earthquake location and phase identification. *Geophysical Journal International*, **105**, 429–465.

Kind, R., 1985. The reflectivity method for different source and receiver structures and comparison with GRF data. *Journal of Geophysics*, **58**, 146–152.

Kind, R., and Vinnik, L. P., 1988. The upper-mantle discontinuities underneath the GRF array from P-to-S converted phases. *Journal of Geophysics*, **62**, 138–147.
Kind, R., Kosarev, G. L., and Petersen, N. V., 1995. Receiver functions at the stations of the German Regional Seismic Network (GRSN). *Geophysical Journal International*, **121**, 191–202.
Kosarev, G., Kind, R., Sobolev, S. V., Yuan, X., Hanka, W., and Oreshin, S., 1999. Seismic evidence for a detached Indian Lithospheric mantle beneath Tibet. *Science*, **283**, 1306–1309.
Kumar, P., Kind, R., and Yuan, X. H., 2010. Receiver function summation without deconvolution. *Geophysical Journal International*, **180**, 1223–1230.
Langston, C. A., 1979. Structure under Mount Rainer, Washington, inferred from teleseismic body waves. *Journal of Geophysical Research*, **84**, 4749–4762.
Langston, C. H. A., and Hammer, J. K., 2001. The vertical component P-wave receiver function. *Bulletin. Seismological Society of America*, **91**(6), 1805–1819.
Levin, V., and Park, J., 1997. P-SH conversions in a flat-layered medium with anisotropy of arbitrary orientation. *Geophysical Journal International*, **131**, 253–266.
Li, X. Q., Yuan, X. H., and Kind, R., 2007. The lithosphere-asthenosphere boundary beneath the western United States. *Geophysical Journal International*, **170**(2), 700–710.
Ligorria, J. P., and Ammon, C. J., 1999. Iterative deconvolution and receiver function estimation. *Bulletin. Seismological Society of America*, **89**, 1395–1400.
Mohsen, A., Hofstetter, R., Bock, G., Kind, R., Weber, M., Wylegalla, K., Rumpker, G., and DESERT Group, 2005. A receiver function study across the Dead Sea Transform. *Geophysical Journal International*, **160**, 948–960.
Montalbetti, J. F., and Kanasewich, E. R., 1970. Enhancement of teleseismic body phases with a polarization filter. *Geophysical Journal International*, **21**(2), 119–129.
Rondenay, S., 2009. Upper mantle imaging with array recordings of converted and scattered teleseismic waves. *Surveys in Geophysics*, **30**, 377–405.
Ryberg, T., and Weber, M., 2000. Receiver function array: a reflection seismic approach. *Geophysical Journal International*, **41**, 1–11.
Schulte-Pelkum, V., Monsalve, G., Sheehan, A., Pandey, M. R., Sapkota, S., Bilham, R., and Wu, F., 2005. Imaging the Indian subcontinent beneath the Himalaya. *Nature*, **435**, 1222–1225, doi:10.1038/nature03678.
Vinnik, L. P., 1977. Detection of waves converted from P to SV in the mantle. *Physics of the Earth and Planetary Interiors*, **15**, 39–45.
Yuan, X., Kind, R., Li, X., and Wang, R., 2006. S receiver functions: synthetics and data example. *Geophysical Journal International*, **175**(2), 555–564.
Yuan, X., Li, X., Wölbern, I., and Kind, R., 2007. Tracing the Hawaiian mantle plume by converted seismic waves. In Ritter, J. R. R., and Christensen, U. R. (eds.), *Mantle Plumes – A Multidisciplinary Approach*. Berlin: Springer, pp. 49–69.
Zhu, L. P., and Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions. *Journal of Geophysical Research*, **150**, 2969–2980.

Cross-references

Body Waves
Deep Seismic Reflection and Refraction Profiling
Earth's Structure, Continental Crust
Energy Partitioning of Seismic Waves
Lithosphere, Continental
Lithosphere, Oceanic
Seismic Phase Names: IASPEI Standard
Seismic, Migration
Seismological Networks

SEISMIC, REFLECTIVITY METHOD

Mrinal K. Sen
Department of Geological Sciences, UT Institute for Geophysics, Jackson School of Geosciences,
The University of Texas at Austin, Austin, TX, USA

Synonyms

Reflection matrix method

Definition

Reflectivity method is a semi-analytic method for computing synthetic seismograms in layered elastic media. The method has now been extended to anisotropic and laterally varying media.

The reflectivity method

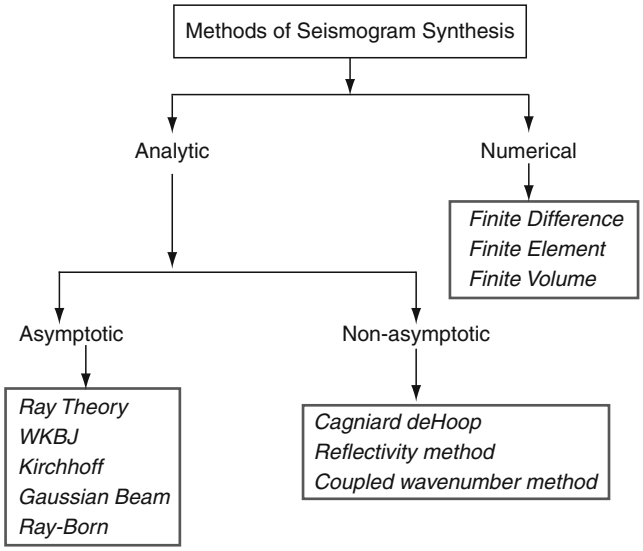
Introduction

Synthetic seismograms are now used routinely for interpretation of field seismic data and evaluating the performance of seismic processing algorithms. Generation of synthetic seismograms essentially involves solution of the partial differential equation (PDE) for seismic wave propagation. Different forms of the PDE are available for different earth models (e.g., acoustic, isotropic, elastic, anisotropic, homogeneous, and laterally inhomogeneous). Brute force numerical methods such as finite differences and finite elements can, in principle, handle models of any complexity. However, they are limited by the speed of computation and computer memory. Even with our latest computer technology, generation of realistic 3D synthetic seismograms is not possible within reasonable clock time.

Much effort has been spent over the years in developing methods that are fast and accurate. We are left with three choices:

- Approximate the earth model and generate analytic or semi-analytic solutions.
- Derive approximate (asymptotic) solution for realistic heterogeneous models.
- Derive pure numerical solution for general heterogeneous models.

Figure 1 provides a fairly complete list of the methods currently available for seismic modeling. Pure numerical methods based on finite-difference (Virieux, 1984) and finite-element (Marfurt, 1984; Komatitsch and Tromp, 1999; DeBasabe and Sen, 2007) approaches generate complete solutions but these methods become prohibitively expensive at high frequencies. The ray-based methods generate asymptotic solutions at infinite frequency and can be applied to models of general complexity as long as rays can be traced through the medium (Cerveny, 2001). Although the original WKBJ method is valid for layered media, its extension to laterally varying media is the Maslov method (Chapman and Drummond, 1982).



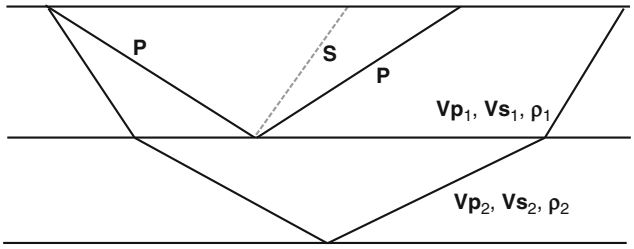
Seismic, Reflectivity Method, Figure 1 Different methods for computing synthetic seismograms: the reflectivity method is a semi-analytic method that is generally valid for layered earth models.

The methods such as the reflectivity (Fuchs and Muller, 1971) or Cagniard-deHoop method (e.g., Aki and Richards, 2000) are included in the category of non-asymptotic analytic method. They were originally proposed for layered (1D) isotropic media and are still commonly used in such media although some extensions of these methods to anisotropy and lateral heterogeneity have indeed been reported. Although I included these methods in the category of "analytic method," their final evaluation is done numerically. Thus strictly speaking, they should be included in the category of "semi-analytic" methods. The speed of computation of these methods can be improved by using asymptotic approximation at the final stage of evaluation of an oscillatory integral.

The primary goal of this entry is to provide an overview of the reflectivity method – perhaps the most popular method of seismogram synthesis. In the following sections, I will summarize the theory and numerical methods pertaining to reflectivity and provide some numerical examples. Finally, I will discuss an extension of the reflectivity method to laterally inhomogeneous media.

Plane waves

The plane waves are fundamental to understanding wave propagation. It has been demonstrated that a point source response can be generated by a weighted sum of individual plane wave responses (e.g., Aki and Richards, 2000). In a homogeneous medium, plane waves propagate without spreading. In a stratified medium, plane waves remain planar during their propagation. However, as they propagate through a stack of flat layers (Figure 2), they undergo changes in amplitude and phase.



Seismic, Reflectivity Method, Figure 2 A stack of layers in which each layer is characterized by its elastic parameters and layer thickness: as an incident plane wave propagates through the stack of layer, it undergoes mode conversion and changes in amplitude due to reflection and transmission.

- The amplitude changes can be accounted for by imposing boundary conditions of continuity of displacement and stress resulting in expressions for the so-called plane wave reflection and transmission coefficients.
- Travel time changes can be computed using individual layer velocities for locally linear ray-paths.
- Changes in propagation angle can be computed by imposing Snell's law of preservation of horizontal slowness.
- In a stratified medium, an incident plane wave generates converted waves and internal multiples with appropriate amplitudes given by reflection and transmission coefficients.

Figure 2 displays a stack of layers in which each layer is characterized by its elastic parameters (densities and, P- and S-wave velocities) and thicknesses in an elastic isotropic medium. An incident plane P-wave characterized by a ray (normal to the plane) given by the angle it makes with the vertical axis is shown in the top layer. At the first layer boundary, it generates a reflected P-wave and a reflected SV-wave (converted). It also generates several reverberations with the top free surface (not shown in the figure). As the plane wave enters the second layer, it changes its propagation direction and similarly generates several converted waves and multiples. The primary task in generating synthetic seismograms is to keep track of all of these paths and account for all these amplitude and phase changes. The reflectivity method does exactly that. However, before we look into the details of the reflectivity method, let us first examine some simple cases with a few rays.

Simple plane wave synthetics

It is fairly straightforward to generate plane wave synthetic seismograms for a layered medium for a few rays or plane waves. Two fundamental parameters involving plane wave seismograms are the ray parameter or horizontal slowness denoted by p and vertical delay time denoted by τ. They are defined as follows

$$p = \frac{\sin\theta}{v}, \quad q = \frac{\cos\theta}{v}, \tag{1}$$

where, θ is the vertical angle of the ray, q is called the vertical slowness and

$$\tau = 2hq, \tag{2}$$

where, h is the layer thickness.

Now referring to Figure 2, the *PP* reflection delay time for layer 1 is

$$\tau_1^{PP} = 2h_1 q_1^P, \tag{3}$$

and the *PS* reflection delay time is

$$\tau_1^{PS} = h_1\left(q_1^P + q_1^s\right). \tag{4}$$

In the above equations, the subscript represents the layer and the superscript represents the wavemode under consideration. The plane wave response of the primary *PP* mode from the interface between layer 1 and layer 2 can be written down in the frequency-slowness domain as

$$R(\omega, p) = R_1^{PP} \exp\left(i\omega 2h_1 q_1^{PP}\right), \tag{5}$$

where, R_1^{PP} is the *PP* plane wave reflection coefficient for interface 1. Similar expression can be written down for the *PS* mode primary reflection.

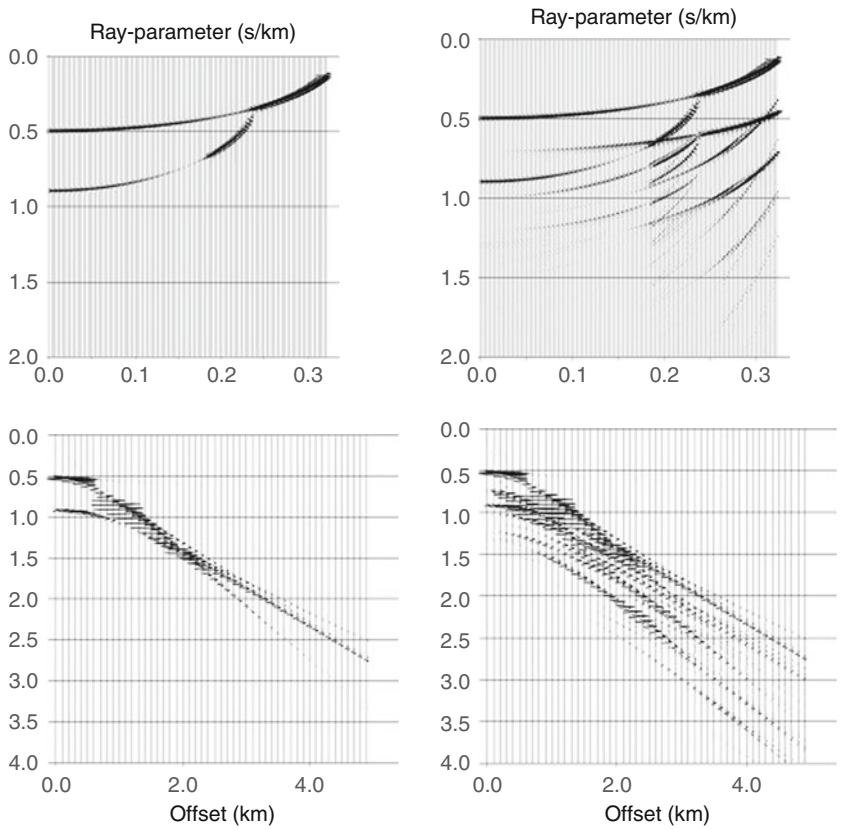
$$R(\omega, p) = R_1^{PS} \exp\left(i\omega h_1\left(q_1^P + q_1^S\right)\right). \tag{6}$$

To include the reflection from the second layer, we simply need to add a term appropriate for the second layer (similar to Equation 3) resulting in the following expression

$$R(\omega, p) = T_{1D}^{PP} R_2^{PP} T_{1U}^{PP} \exp\left(i\omega\left(2h_1 q_1^{PP} + 2h_2 q_2^{PP}\right)\right), \tag{7}$$

where we include the effects of transmission using downgoing (with subscript D) and upgoing (with subscript U) transmission coefficients. Expressions similar to Equation 7 can also be written down for mode-converted waves.

Examples of plane wave and point source synthetic seismograms are shown in Figure 3; these are generated for a vertical point force and only the traces of vertical component of displacement are shown. The upper panel



Seismic, Reflectivity Method, Figure 3 Examples of plane wave and point source synthetic seismograms for a model comprising two elastic layers over a half space: these were generated for a vertical point force and only the traces of vertical component of displacement are shown. The upper panel shows the plane wave seismograms in τ-p domain while the lower panel shows the corresponding x-t domain synthetic seismograms. The left panel shows seismograms for primary *PP* reflections only. The right panel shows the seismograms that include primary *PP*, *PS* and internal multiples from the second layer.

shows the plane wave seismograms in τ-p domain while the lower panel shows the corresponding x-t domain synthetic seismograms. The left panel shows seismograms for primary *PP* reflections only. Note two distinct reflections from the two layer boundaries. The right panel shows seismograms that include primary *PP*, *PS,* and internal multiples from the second layer.

It is fairly straightforward to compute seismograms for a few selected phases for a few layers. In reality, we have many layers and an infinite number of ray-paths – these are not easy to count. The reflectivity method offers an elegant algorithm for computing full response without having to count rays. Of course, we may choose to compute a few selected phases under reflectivity formulation as well, if desired.

Theory

Detailed developments of the theory of the reflectivity method for isotropic media can be found in Fuchs and Muller (1971) and Kennett (1983). Derivations for an anisotropic layered medium can be found in Booth and Crampin (1983) and Fryer and Frazer (1984). We start with the two fundamental equations in seismology, namely, the constitutive relation and the equation of motion given by,

$$\tau = \mathbf{C} : \nabla \mathbf{u} \tag{8}$$

$$\rho\omega^2 \mathbf{u} = \nabla \cdot \tau + \mathbf{f}, \tag{9}$$

where, τ is a second rank stress tensor, **C** is the fourth rank elastic stiffness tensor, **u** is the displacement vector, ω is the angular frequency, and **f** is a body force term. Note that the above equations are in frequency domain.

The fundamental assumption in reflectivity is that the medium is layered and therefore, we can apply transformations over x and y coordinates to horizontal wave numbers K_x and K_y or horizontal slowness p_x and p_y. For isotropic and layered transversely isotropic media, we can make use of cylindrical symmetry and transform to radial ray parameter p and azimuth ϕ. For a general anisotropic medium, we simply apply Fourier transforms over x and y coordinates to transform Equations 8 and 9 into horizontal wave numbers or ray-parameter domain. Once that is achieved, all the analytic developments are done in the frequency ray-parameter domain.

The equation of motion and the constitutive relation are thus transformed into the following system of ordinary differential equations in depth z, by applying a Fourier transform in x and y

$$\mathbf{b} = i\omega\, \mathbf{A}\mathbf{b} + \mathbf{f}, \tag{10}$$

where, $\mathbf{b} = \left[u_x\, u_z\, u_y\, \tau_{xz}\, \tau_{zz}\, \tau_{yz}\right]^T = \mathbf{b}(\omega\,, \mathbf{p})$is the stress-displacement vector which is a function of frequency ω and horizontal slowness p, **A**(ω, **p**) is the system matrix that is a function of elastic coefficients, and **f** is a body force term. For isotropic and transversely isotropic media, the Equation 10 decouples into two systems, namely a P-SV (4 × 4) system and an SH (2 × 2) system. The solution of the ODE (Equation 10) can be carried out by a propagator matrix method (Gilbert and Backus, 1966). It is well known that the propagator matrix is generally unstable due to growing exponentials (e.g., Jensen et al., 1993). Stable solutions can be obtained by one of the three methods:

- A global matrix approach (Schmidt and Tango, 1986).
- Compound matrix approach where we define a new system of ODE in which the elements of the new system matrix are the minors of the original system matrix. The original 4 × 4 P-SV system maps into a 6 × 6 system (Dunkin, 1965; Phinney et al., 1987).
- An invariant imbedding or a reflection matrix approach (Kennett, 1983).

Of the three methods listed above, the compound matrix and the reflectivity methods have been widely used in seismology. Table 1 summarizes the two methods. The reflectivity method has been very popular because of its ray-interpretation and easy generalization to azimuthally anisotropic media. In the unconditionally stable reflection matrix approach (Kennett, 1983), the propagation uses the eigenvalues and eigenvectors of the system matrix **A**; the eigenvalues are the vertical phase functions. These

Seismic, Reflectivity Method, Table 1 Reflection matrices or compound matrices?

Reflectivity approach	Compound matrix approach
Unconditionally stable	**Unconditionally stable**
For the P-SV case, the algorithm evaluates four 2 × 2 complex reflection/transmission matrices and iteration equations for each frequency and ray-parameter.	For the P-SV case, the algorithm has fewer floating-point operations than the reflectivity method. The method is computationally faster than the reflectivity method.
Derivatives or differential seismograms can be computed analytically for isotropic and vertically transversely isotropic (VTI) media	Derivatives or differential seismograms can be computed analytically for isotropic and VTI media
Derivatives can be computed by semi-analytic approach for general anisotropic media	–
Requires VERY CAREFUL coding	**Coding is very EASY**
Most efficient for general anisotropic media	• Very slow for general anisotropy
Ray-interpretation is possible. Selected rays and selected number of multiples can be included. The algorithm is VERY FLEXIBLE	Complete solution is derived

eigenvalues and eigenvectors are used to define four upgoing and downgoing reflection and transmission coefficient matrices, $\mathbf{R}_D$, $\mathbf{T}_D$, $\mathbf{R}_U$, and $\mathbf{T}_U$. These are propagated through the stack of layers to obtain a composite reflection matrix that includes the effects of reflection, transmission, mode conversion, and internal multiples. Kennett (1983) derived the following iteration equation (Kennett, 1983; p.127), which can be used to compute the R/T matrices of a zone AC when those of zones AB and BC are known:

$$\begin{aligned}
\mathbf{R}_D^{AC} &= \mathbf{R}_D^{AB} + \mathbf{T}_U^{AB}\mathbf{R}_D^{BC}\left[\mathbf{I} - \mathbf{R}_U^{AB}\,\mathbf{R}_D^{BC}\right]^{-1}\mathbf{T}_D^{AB} \\
\mathbf{T}_D^{AC} &= \mathbf{T}_D^{BC}\left[\mathbf{I} - \mathbf{R}_U^{AB}\,\mathbf{R}_D^{BC}\right]^{-1}\mathbf{T}_D^{AB} \\
\mathbf{R}_U^{AC} &= \mathbf{R}_U^{BC} + \mathbf{T}_D^{BC}\mathbf{R}_U^{AB}\left[\mathbf{I} - \mathbf{R}_D^{BC}\,\mathbf{R}_U^{AB}\right]^{-1}\mathbf{T}_U^{BC} \\
\mathbf{T}_U^{AC} &= \mathbf{T}_U^{AB}\left[\mathbf{I} - \mathbf{R}_D^{BC}\,\mathbf{R}_U^{AB}\right]^{-1}\mathbf{T}_U^{BC}.
\end{aligned} \tag{11}$$

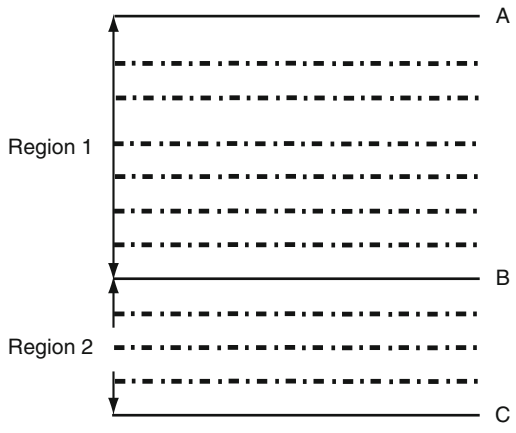
Equation 11 is the most fundamental development of the reflectivity formulation. Given the R/T matrices (which include layer propagation and interaction terms) of two consecutive regions AB and AC (Figure 4), we can compute the response of region AC using Equation 11. Thus, we can propagate through the stack of layers resulting in an unconditionally stable algorithm.

Let us examine the downward reflection matrix in Equation 11 in slightly more detail. Using the matrix identity

$$(\mathbf{I} - \mathbf{A})^{-1} = I + \mathbf{A}\mathbf{A} + \mathbf{A}\mathbf{A}\mathbf{A} + \ldots, \tag{12}$$

we can write

$$\begin{aligned}
\mathbf{R}_D^{AC} &= \mathbf{R}_D^{AB} + \mathbf{T}_U^{AB}\mathbf{R}_D^{BC}\left[\mathbf{I} - \mathbf{R}_U^{AB}\,\mathbf{R}_D^{BC}\right]^{-1}\mathbf{T}_D^{AB} \\
&= \mathbf{R}_D^{AB} + \mathbf{T}_U^{AB}\mathbf{R}_D^{BC}\mathbf{T}_D^{AB} + \mathbf{T}_U^{AB}\mathbf{R}_D^{BC}\mathbf{R}_U^{AB}\,\mathbf{R}_D^{BC}\mathbf{T}_D^{AB} + \ldots
\end{aligned} \tag{13}$$



Seismic, Reflectivity Method, Figure 4 Zones AB and BC from a stack of layers; if the upgoing and downgoing reflection/transmission matrices for these two zones are known, those for the entire stack *AC* can be generated using the iteration equations.

Figure 5 clearly shows that Equation 13 includes all the internal multiples and mode-converted waves. We may also choose to retain a few terms in the expansion and thus compute exact response of a few rays.

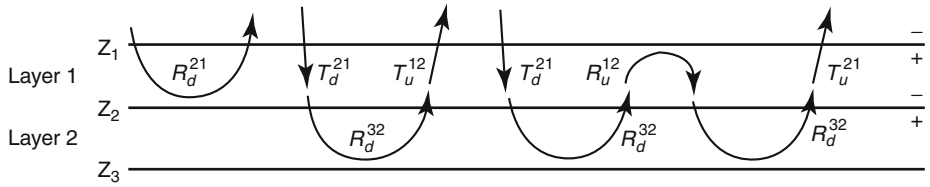
While computing the up/downgoing reflection/transmission matrices through the stack of layers, we need to save the appropriate matrices through the source, receiver layers, and the free surface. Finally, one can compute the entire stress/displacement vector comprising the components of displacement and traction at any receiver location.

Note that we obtain our solution in the frequency ray-parameter domain. An inverse temporal Fourier transforms results in (τ-p) seismograms. Synthetics in the offset-time domain can be obtained by plane wave synthesis of the (ω, p) or (τ-p) seismograms. A general flow chart for computing synthetic seismograms by the reflectivity method is shown in Figure 6. Note that the ray-parameter, layer and frequency loops can be interchanged depending on the application. As stated earlier, intermediate results in (τ-p) domain are very useful for understanding wave propagation and amplitude effects since they are devoid of spherical spreading loss (Fryer, 1980).

Computational issues

It has been demonstrated that the reflectivity algorithm reviewed in this entry is unconditionally stable (Kennett, 1983). However, one must incorporate some practical strategies in computing noise-free synthetic seismograms (e.g., Mallick and Frazer, 1987).

- *Spatial aliasing*: Generation of offset synthetics from plane wave seismograms requires evaluation of an oscillatory integral. The integrand becomes more and more oscillatory with increase in offset and frequency. Thus a practical strategy would be to use a frequency dependent sampling interval in ray-parameters. Alternatively, one can use the minimum ray-parameter increment appropriate for the highest frequency and the largest offset of interest. We also need to use smooth tapers at the high and low ends of the ray-parameter window to avoid truncation phases.
- *Temporal aliasing*: It can be avoided either by using a very large time window or adding a small imaginary part in the frequency following a method outlined in Phinney (1965).
- *Parallelization*: For high frequency and large offset calculation, the reflectivity calculation may be computationally intensive especially for azimuthally anisotropic media, where reflectivity calculation needs to be performed for a large number of ray parameters. Note that (Figure 6) almost the entire reflectivity computation can be done in parallel resulting in an algorithm that may be termed "embarrassingly parallel." One of the simplest ways to parallelize is to distribute the computation of ray-parameter traces equally to the available nodes of a computer (e.g., Roy et al., 2005). The master node collects all the ray parameter traces (note that there is no communication in between the



Seismic, Reflectivity Method, Figure 5 A demonstration of how the iteration equation includes all the reflection and transmission effects.

General Flow chart

- Loop over ray-parameter
  - Loop overe layer
    - Compute R/T (up going and downgoing) for all the interfaces
    - loop over frequency
      - evaluate iteration equation
    - end loop over frequency
  - end loop over layer
  - inverse FFT for (tau-p) seismograms
- End loop over ray-parameter
- Sum the plane wave responses to compute point source synthetics

Seismic, Reflectivity Method, Figure 6 A general flow chart for developing a reflectivity code.

nodes required for the calculation of ray parameter trace). This results in an almost linear speedup as a function of the number of nodes. MPI (message passing interface) can be used for parallelization; the reflectivity code can also be parallelized using simple compiler directives, if coded carefully.

- *Attenuation*: In order to include attenuation, we need to specify Q values for P- and S- waves for each layer. For frequency dependent attenuation, the velocities are complex and dependent on frequency. In such cases, the layer loop (Figure 6) becomes the deepest loop.

Applications

Exploration scale modeling: Numerous applications of reflectivity synthetic seismograms for exploration can be found in literature. This algorithm is often used to generate synthetic seismograms from well logs and then used in calibrating a seismic gather at the well location and well tie (e.g., Sen, 2006). Figure 7a shows α_0, β_0, density, and VTI anisotropy parameters ε and δ derived from a well log as a function of two way time. The plane wave synthetics for an isotropic and VTI model and their differences are shown in Figure 7b. Note that for isotropic calculations, ε and δ values were set equal to zero; for VTI calculations, P- and S-wave velocities from well logs were treated as vertical P and S velocities respectively. Offset synthetics for isotropic and VTI models and their differences are shown in Figure 7c. As expected, the effect of anisotropy is more pronounced at large offsets or ray-parameters.

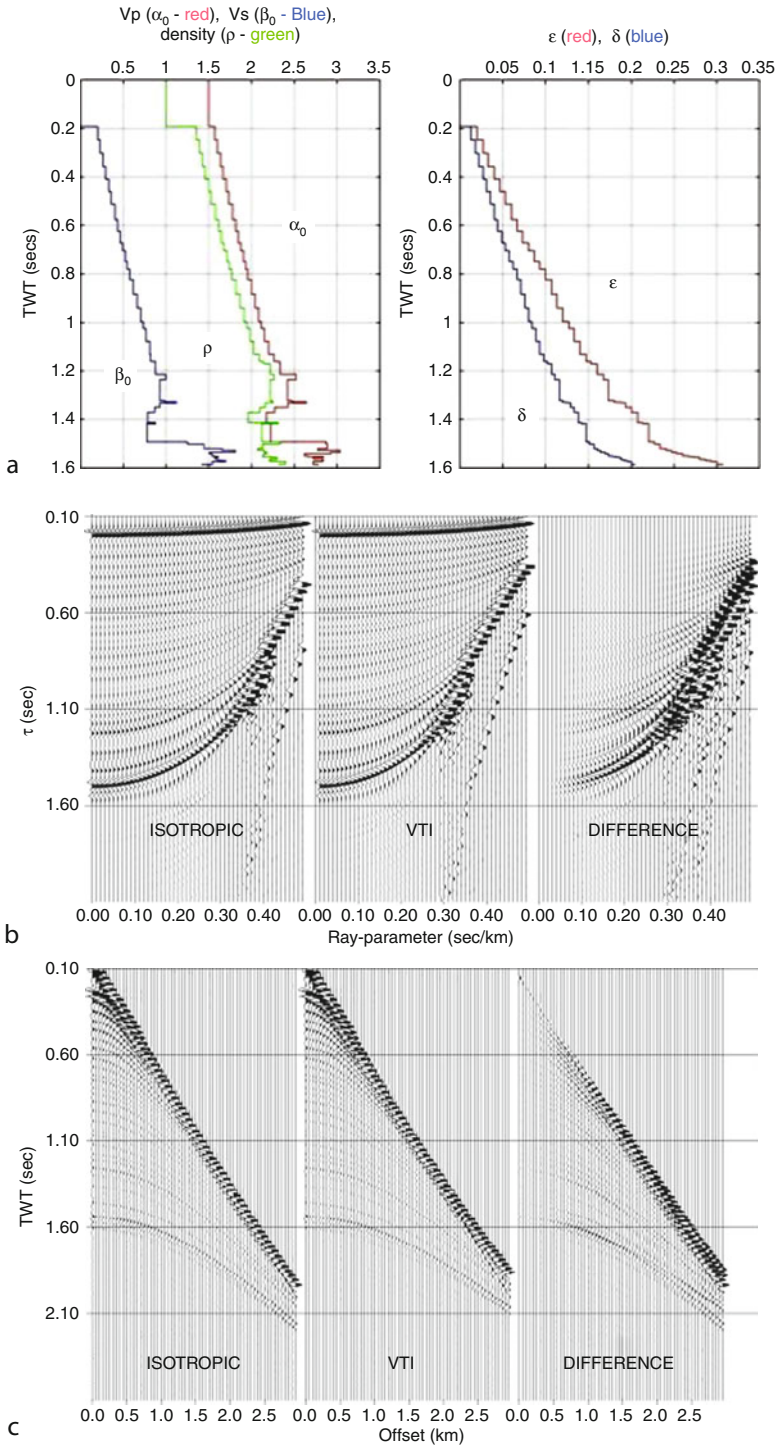
Modeling anisotropy in the core–mantle transition zone (D″): For modeling at regional and teleseismic distances, we need to apply earth-flattening transformation to the models prior to computing synthetic seismograms using a reflectivity code. An example of such modeling is given in Pulliam and Sen (1998) in which attempts were made to model shear wave splitting observations of S-wave phases that propagate nearly horizontally through the core–mantle transition zone. Figure 8 shows the data and the corresponding synthetic seismograms for a station at Hockley from earthquakes from Tonga-Fiji. The model used for reflectivity calculation includes transversely isotropic layers with a horizontal axis of symmetry within the D″ zone.

Modeling in azimuthally anisotropic media (comparison with FD): An example of comparison of azimuthally anisotropic reflectivity synthetic seismograms with those computed by a 3D azimuthally anisotropic finite-difference code is shown in Figure 9 (Bansal and Sen, 2008). The synthetics computed by the two methods are in excellent agreement.

Extension to laterally heterogeneous media

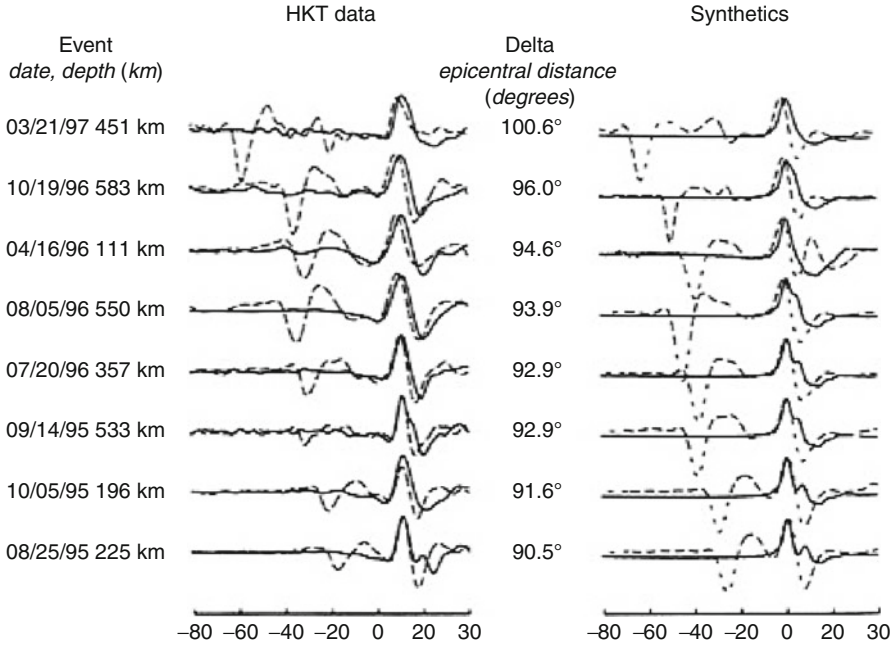
Extension of the reflectivity method to two dimensions, where the velocities and densities are allowed to vary laterally as well, is nontrivial. Koketsu et al. (1991) and Sen and Pal (2009) developed extended reflectivity methods for the case of homogeneous layers separated by irregular or curved interfaces which do not cross each other. Lateral media variations within a layer cannot be easily taken into account in a reflectivity-type formalism because the concept of an interface is inherent in it except using the concepts of pseudo-differential operators and Fourier Integral operators (McCoy et al., 1986).

Across a flat surface, the horizontal slowness p is conserved according to Snell's law. However, across an irregular interface, an incident ray with a particular p is scattered into different plane waves with slownesses p'. This is mathematically equivalent to Fourier transforming the dependence of z on x to the scattered set of p' or k' (wave number). Now the coefficients $\mathbf{R_D}(p)$, $\mathbf{R_U}(p)$, $\mathbf{T_D}(p)$, and $\mathbf{T_U}(p)$ become functions of both p and p'. The 2D extension of the method as proposed by Koketsu et al. (1991) involves explicit evaluation of boundary conditions that results in a matrix formulation involving several matrix inversions in the coupled ray-parameter domain; numerous numerical artifacts are caused by such matrix operations. An alternate asymptotic approach was

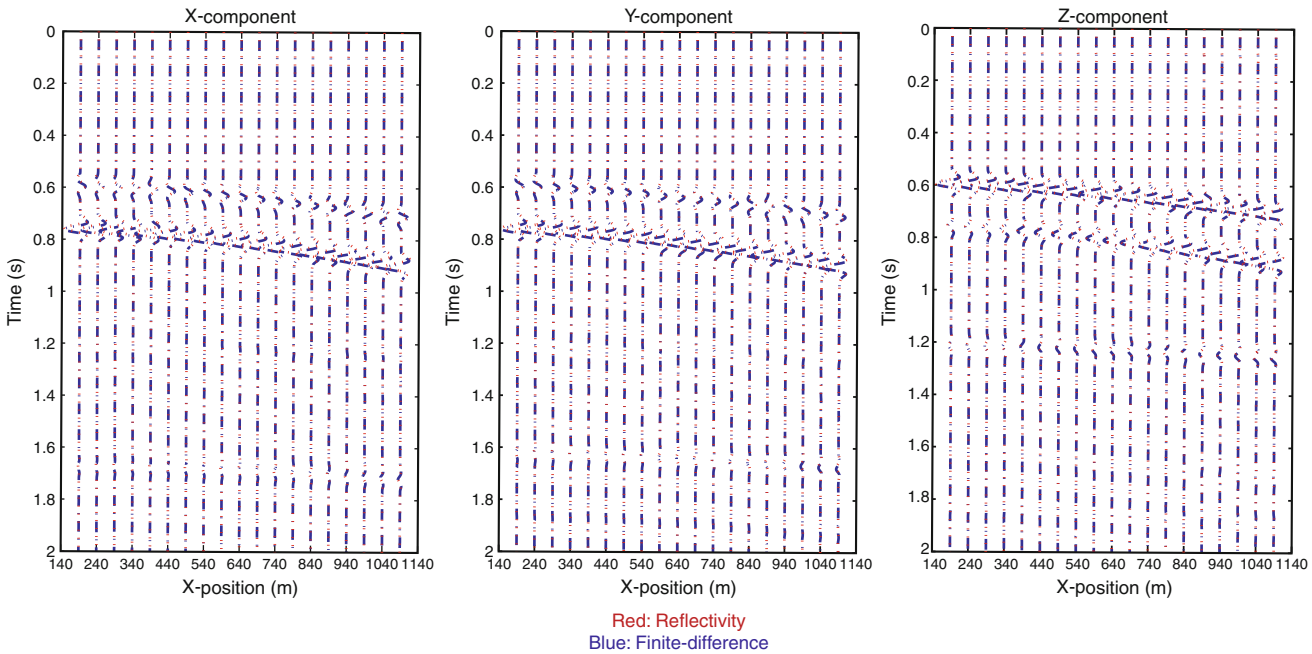


Seismic, Reflectivity Method, Figure 7 (**a**) *Vp*, *Vs*, density, *e* and *d* for a well log model; (**b**) plane wave synthetic seismograms for an isotropic model (*left*), VTI model (*middle*), and the difference (*right*); (**c**) offset synthetic seismograms for an isotropic model (*left*), VTI model (*middle*), and the difference (*right*).

proposed by Sen and Pal (2009) by examining a tangent-plane or Kirchhoff formulation in the plane wave domain to replace the exact boundary condition evaluation (Sen and Frazer, 1991). All other reflectivity operations such as the invariant imbedding or iterative computation of reflection, transmission, multiple, and mode conversions can be readily applied even under this approximation. This new algorithm computes noise-free seismograms even



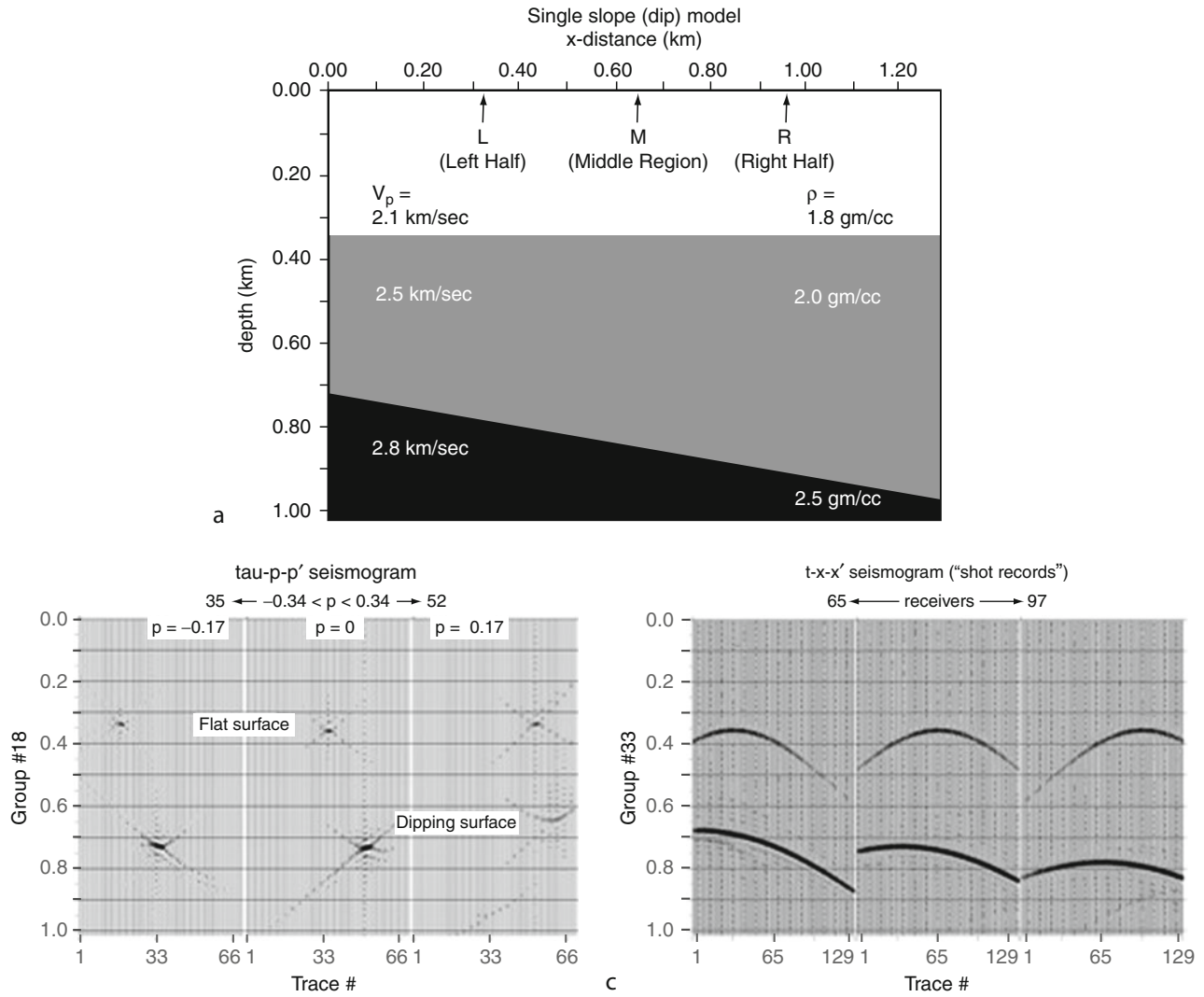
Seismic, Reflectivity Method, Figure 8 Comparison of S-wave arrivals from Tonga-Fiji events recorded at Hockley station. Synthetic seismograms (*right panel*) were computed using a reflectivity code for a model containing transversely isotropic layers with a horizontal axis of symmetry within the core–mantle transition zone. Solid lines represent transverse and dashed lines represent radial component seismograms. Reproduced with permission from Geophysical Journal International.



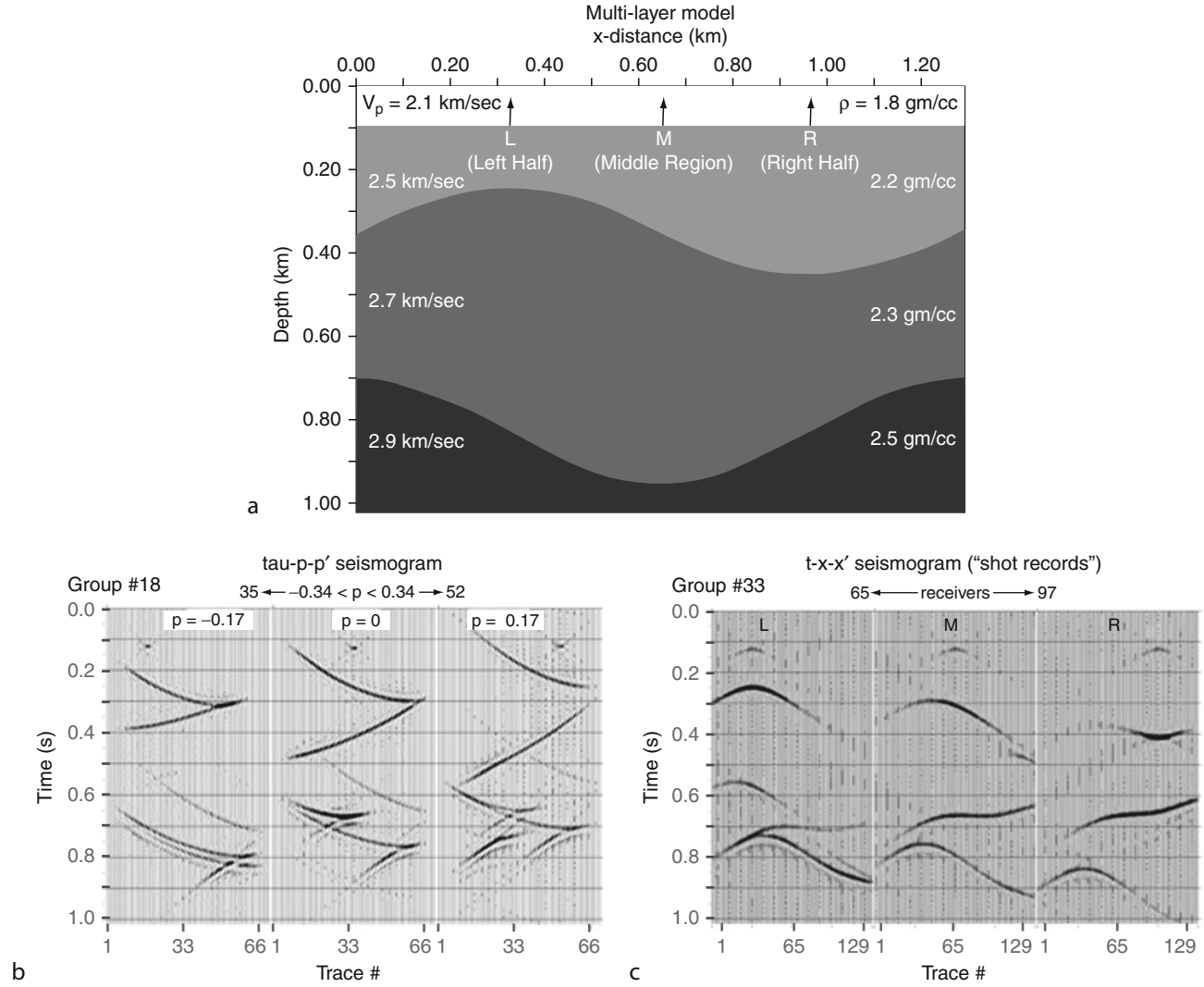
Seismic, Reflectivity Method, Figure 9 A comparison between reflectivity and 3D finite-difference synthetic seismograms for a two-layer azimuthally anisotropic earth model. Red: reflectivity; Blue: finite difference. Reproduced with permission from Geophysical Prospecting.

with coarse sampling of the interfaces and ray-parameters. Approximate calculation of reflection/transmission coefficients, however, does not include multiple interaction of a plane wave with an interface.

Sen and pal (2009) performed computations in the coupled slowness domain and thus, multiple shot-receiver data can be synthesized rapidly. Intermediate results in the coupled slowness space provide important insight into understanding wave propagation in heterogeneous media. Figure 10 shows synthetic seismograms for a model with two homogeneous layers over a half space (Figure 10a) such that the top interface is flat and the second interface is dipping with a constant dip. Figure 10b shows synthetic seismograms in (τ-p-p') domain; note that for the flat layer, the scattered p is the same as the incident ray-parameter and the same for the dipping reflector is different from the incident ray-parameter. Synthetic shot gathers for three shot locations shown in Figure 10a are displayed in Figure 10c. These were all computed using a 2D reflectivity code in which internal multiples were not included. Similarly, Figure 11 shows the synthetic seismograms for a 2 layer over half space model where both the interfaces are curved. Note the triplicated arrivals in the seismograms in both τ-p-p' domain (Figure 11b) and shot gathers (Figure 11c). Note the triplicated arrivals in the seismograms due to curved interfaces.



Seismic, Reflectivity Method, Figure 10 (**a**) A model with two homogeneous layers over a half space; (**b**) τ-p-p' synthetic seismograms for the model shown in (**a**): note the shift in p value for the reflection from the dipping interface; (**c**) shot gathers for the mode shown in (**a**) – asymmetry in the reflection pattern from the second interface is modeled very well. Reproduced with permission from Geophysical Journal International.



Seismic, Reflectivity Method, Figure 11 (**a**) A model with two homogeneous layers over a half space; (**b**) τ-*p*-*p*′ synthetic seismograms for the model shown in (**a**): note the shift in *p* values and triplications for the reflection from the curved interfaces; (**c**) shot gathers for the mode shown in (**a**) – triplications in the reflections modeled very well. Reproduced with permission from Geophysical Journal International.

Summary

Seismic modeling is crucial to understanding seismic wave propagation and interpreting field seismic records. Several methods for seismic modeling are available; they vary in accuracy, model assumptions, and computational speed. Of these, the reflectivity method is perhaps the most popular. The original reflectivity formulation is valid for 1D elastic media. It is capable of computing a complete response of a layered earth model. If needed, this can also be designed to compute responses of a few selected phases. The cost of computation grows with the number of layers, increase in frequency, and offset. Most isotropic models can now be computed fairly rapidly on standard desktop workstations or personal computers. Azimuthally anisotropic models require computations in two wave numbers and therefore can be computationally demanding. Parallel algorithms can be easily developed for rapid computation of reflectivity synthetic seismograms for large models. For 1D full waveform inversion, the reflectivity forward modeling is invoked a large number of times where a parallel reflectivity algorithm is also essential. The reflectivity algorithm has been used in a wide variety of applications including seismic modeling in exploration and whole earth scales. Full waveform inversion methods based on reflectivity forward modeling are becoming increasingly popular (e.g., Sen and Roy, 2003). The method has also been extended to modeling in laterally heterogeneous media and to electromagnetic modeling in layered media (Sena et al., 2008).

Bibliography

Aki, K., and Richards, P. G., 2000. *Quantitative Seismology*. San Francisco: W. H. Freeman.

Bansal, R., and Sen, M. K., 2008. Finite difference of S-wave splitting in anisotropic media. *Geophysical Prospecting*, **56**, 293–312.

Booth, D. C., and Crampin, S., 1983. The anisotropic reflectivity technique: theory. *Geophysical Journal of the Royal Astronomical Society*, **72**(3), 755–766.

Cerveny, V., 2001. *Seismic Ray Theory*. Cambridge: Cambridge University Press.

Chapman, C. H., and Drummond, R., 1982. Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory. *Bulletin. Seismological Society of America*, **72**, S277–S317.

DeBasabe, J., and Sen, M. K., 2007. Grid dispersion and stability criteria of some finite element methods for seismic wave propagation modeling. *Geophysics*, **72**(6), T81–T95.

Dunkin, J. W., 1965. Computation of modal solutions in layered elastic media at high frequencies. *Bulletin. Seismological Society of America*, **55**, 335–338.

Fryer, G. J., 1980. A slowness approach to the reflectivity method of seismogram synthesis. *Geophysical Journal of the Royal Astronomical Society*, **63**(3), 747–758.

Fryer, G. J., and Frazer, L. N., 1984. Seismic waves in stratified anisotropic media. *Geophysical Journal International*, **78**(3), 697–710.

Fuchs, K., and Mueller, G., 1971. Computation of synthetic seismograms with the reflectivity method and comparison with observations. *Geophysical Journal of the Royal Astronomical Society*, **23**, 417–433.

Gilbert, F., and Backus, G. E., 1966. Propagator matrices in elastic wave and vibration problems:. *Geophysics*, **31**, 326–332.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H., 1993. *Computational Ocean Acoustics*. New York: AIP.

Kennett, B. L. N., 1983. *Seismic Wave Propagation in Stratified Media*. Cambridge: Cambridge University Press.

Koketsu, K., Kennett, B. L. N., and Takenaka, H., 1991. 2-D reflectivity method and synthetic seismograms for irregularly layered structures – II. Invariant embedding approach. *Geophysical Journal International*, **105**, 119–130.

Komatitsch, D., and Tromp, J., 1999. Introduction to the spectral element for three dimensional seismic wave propagation. *Geophysical Journal International*, **139**, 806–822.

Mallick, S., and Frazer, L. N., 1987. Practical aspects of reflectivity modeling. *Geophysics*, **52**, 1355–1364.

Marfurt, K. J., 1984. Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, **49**, 533–549.

McCoy, J. J., Fishman, L., and Frazer, L. N., 1986. Reflection and transmission at an interface separating transversely inhomogeneous acoustic half spaces. *Geophysical Journal of the Royal Astronomical Society*, **85**, 543–562.

Phinney, R. A., 1965. Theoretical calculation of spectra of first arrivals in layered elastic mediums. *Journal of Geophysical Research*, **70**, 5107–5123.

Phinney, R. A., Odom, R. I., and Fryer, G. J., 1987. Rapid generation of synthetic seismograms in layered media by vectorization of the algorithm. *Bulletin. Seismological Society of America*, **77**, 2218–2226.

Pulliam, J., and Sen, M. K., 1998. Anisotropy in the core-mantle transition zone may indicate chemical heterogeneity. *Geophysical Journal International*, **135**, 113–128.

Roy, I. G., Sen, M. K., and Torres-verdin, C., 2005. Full waveform inversion using a distributed system of computers. *Concurrency and Computation: Practice and Experience*, **17**, 1635–1685.

Scmidt, H., and Tango, G., 1986. Efficient global matrix approach to the computation synthetic seismograms. *Geophysical Journal of the Royal Astronomical Society*, **84**, 331–359.

Sen, M. K., 2006. *Seismic Inversion*. USA: Society of Petroleum Engineers.

Sen, M. K., and Frazer, L. N., 1991. Multifold phase space path integral synthetic seismograms. *Geophysical Journal International*, **104**, 479–487.

Sen, M. K., and Pal, A., 2009. A reflectivity method for laterally varying media. *Geophysical Journal International*, **178**(2), 792–812.

Sen, M. K., and Roy, I. G., 2003. Computation of differential seismograms and iteration adaptive regularization in pre-stack seismic inversion. *Geophysics*, **68**(6), 2026–2039.

Sena, A., Sen, M. K., and Stoffa, P. L., 2008. Modeling of ground penetrating radar data in stratified media using the reflectivity technique. *Journal of Geophysics and Engineering*, **5**, 129–146.

Virieux, J., 1984. SH-wave propagation in heterogeneous media: velocity-stress finite-difference method. *Geophysics*, **49**, 1933–1957.

Cross-references

Numerical Methods, Finite Difference
Numerical Methods, Finite Element
Seismic Waves, Scattering
Seismic, Ray Theory
Seismic, Waveform Modeling and Tomography

SEISMIC, VISCOELASTIC ATTENUATION

Vernon F. Cormier
Physics Department, University of Connecticut, Storrs, CT, USA

Synonyms

Seismic intrinsic attenuation

Definition

Linear viscoelastic attenuation. The fractional loss of seismic energy in a material in which elastic deformation (strain) induced by one cycle of a seismic wave or mode lags in time the applied stress associated with the wave or mode.

Apparent seismic attenuation. The loss of energy in a propagating seismic wave or standing mode due to viscoelasticity combined with the loss of scattered energy redistributed in time and space by heterogeneity.

Introduction

The amplitude of seismic waves decreases with increasing distance from earthquake, explosion, and impact sources. How this amplitude decreases, how rapidly it occurs, and how it depends on frequency of the seismic waves is fundamentally important to the efforts to describe Earth structure and seismic sources. The decay of amplitude of seismic waves with increasing distance of propagation through earth is known as seismic wave attenuation. The attenuation occurring under high-temperature rheological

conditions in the earth's interior can be called seismic viscoelastic attenuation.

Seismic attenuation and its variation with location within the Earth are useful for determining the anelastic properties of the earth as a function of depth. Seismic attenuation also shows large lateral variations that can be related to lateral variations in geological and geophysical properties not as easily detected by measurement of seismic velocities. In addition to providing information on a physical property, research in seismic attenuation has also been strongly motivated by more practical problems. One problem has been the prediction of ground motion due to probable earthquakes in different regions. The frequency content and decay with distance of this strong ground motion is an important input to the design of earthquake resistant structures and to disaster forecasting (see *Earthquakes, Strong-Ground Motion*). Another problem has been to estimate the size and detectability of underground nuclear tests (see *Seismic Monitoring of Nuclear Explosions*).

How do seismic waves attenuate?

The attenuation of seismic waves is due to three effects: geometric spreading, intrinsic attenuation, and scattering.

Geometric spreading

Geometric spreading leads to an energy density decrease that occurs as an elastic wave front expands with increasing distance from its source. In a homogeneous Earth of constant velocity and density, the geometric spreading of a seismic body wave is proportional to the reciprocal of the distance between source and receiver. In the real Earth, velocity and density vary strongly with depth and less so laterally. Given a model of this variation, however, the geometric spreading of a body wave can be easily calculated (see *Seismic, Ray Theory*).

Intrinsic viscoelastic attenuation

Intrinsic (viscoelastic) attenuation occurs at high temperatures due to internal friction during the passage of an elastic wave. It is controlled by the thermal and defect properties of the medium in which the wave is propagating. It can result in a phase lag between strain and stress giving rise to strain energy dissipation and associated frequency dependence (dispersion) of the relevant modulus or speed of the propagating elastic wave. The microscopic mechanisms of intrinsic attenuation have been described in several different ways, including the resistive and viscous properties of oscillator models of the atoms in crystalline lattices, the movement of interstitial fluids between grain boundaries and cracks (O'Connell and Budiansky, 1977), and the frictional sliding of cracks. Jackson (1993, 2007) reviews laboratory experiments that investigate microscopic mechanisms of intrinsic attenuation. This article concentrates on the measurement of intrinsic attenuation from recordings of seismic waves at great distance.

Scattering attenuation

Scattering attenuation occurs when elastic energy is scattered and redistributed into directions away from the receiver or into waves arriving in later time windows at the receiver (see *Seismic Waves, Scattering*). Scattering takes place by reflection, refraction, and mode conversion of elastic energy by wavelength-scale irregularities in the medium. These irregularities are discontinuous or rapid variations in the velocity and/or density of the medium. In the crust and uppermost mantle, variations in velocity and density can be particularly strong in the lateral as well as the vertical direction.

Linear viscoelasticity

Rheology

A stress is a vector force per unit area applied to a solid. A strain is non-dimensional measure of the deformation of the solid due to the applied stress, such as the change in a length element divided by the original length. The equation that relates stress and strain is sometimes termed the *rheology* or the *constitutive* relation (see *Mantle Viscosity*). A linear viscoelastic rheology can be described by a linear differential equation:

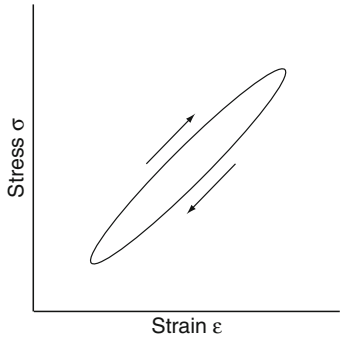
$$L_1\sigma(t) = L_2\varepsilon(t) \tag{1}$$

Where L_1 and L_2 are any linear combinations of operators of the time $\frac{d^n}{dt^n}$ or $\int dt^n$. This type of equation can describe both the elastic strain of a material over a short time interval of applied stress as well as its viscous behavior and flow over a longer time interval (Gross, 1953; Nowick and Berry, 1972; Jackson et al., 2005; Kohistedt, 2007).

Anelastic hysteresis

Seismic oscillations at distances beyond several fault lengths from an earthquake excite small strains less than 10^{-6}. These strains are recoverable during a cycle of seismic oscillation and lag the applied stress of the oscillation in time. Because of the time lag, a cycle of increasing and decreasing stress does not produce a perfectly proportional increase and decrease in strain. Instead a hysteresis loop occurs (Figure 1). The area enclosed by the hysteresis loop is a measure of the energy lost due to heat and internal friction. During the stress cycle associated with the passage of a seismic wave, the energy lost to this internal friction is not available to deform the adjacent regions of the solid ahead of the wave front and therefore the amplitude of the wave decreases.

From the hysteresis curve, one can see that the stress–strain relation cannot be described by a simple constant of proportionality in the time domain. A more complicated relation involving an integral over time is required to describe strain at any instant of time as a function of the prior time history of the applied stress. By Fourier transforming the rheologic equation, however, and keeping only terms describing the short-term anelastic behavior, the stress–strain relation can be simply expressed



Seismic, Viscoelastic Attenuation, Figure 1 Stress–strain hysteresis curve showing the behavior of strain during a cycle of applied stress induced by a propagating seismic wave.

by means of either a complex elastic modulus $\hat{G}(\omega)$ or by its reciprocal, the complex elastic compliance, $\hat{J}(\omega)$:

$$\hat{\sigma}(\omega) = \hat{G}(\omega)\varepsilon(\omega) \tag{2a}$$

$$\hat{\varepsilon}(\omega) = \hat{J}(\omega)\hat{\sigma}(\omega) \tag{2b}$$

The elastic modulus $\hat{G}$ and compliance $\hat{J}$ must be a complex numbers to describe the phase lag of strain. $\hat{G}$ and $\hat{J}$ must also be frequency dependent because the phase lag of strain depends on the time history of stress, the shape of the hysteresis curve changing with different load histories. All the usual measures of anelasticity, including the frequency-dependent quality factor $Q(\omega)$ and the dispersion of the complex phase velocity $\hat{v}(\omega)$ can be expressed in terms of the complex compliance $\hat{J}(\omega)$ (Jackson and Anderson, 1970).The trend of the frequency dependencies can be inferred from the time lag of strain from applied stress.

A feature of the complex modulus is that its real part will be smaller at zero or very low frequency and larger at infinite or very high frequency. That is, there will be an instantaneous response of strain to the applied stress, which is smaller than the eventual equilibrium response after longer time. The difference between the modulus at infinite frequency $G(\infty)$, representing the instantaneous or *unrelaxed* response, and the low-frequency limit of the modulus $G(0)$, for the equilibrium or *relaxed* response, is called the *modulus defec*t ΔG, with

$$\Delta G = G(\infty) - G(0) \tag{3}$$

The relaxed and unrelaxed moduli are pure real numbers that can be determined by observing a sequence of hysteresis curves for increasing frequencies of monochromatic loads. The frequency dependence of the real part of the modulus G at frequencies between 0 and ∞ implies that the propagation of a stress pulse will be dispersive, with higher frequencies traveling faster than lower frequencies.

Q and complex velocity

Since simple mechanical systems, composed of springs and dashpots and simple electric circuits also obey linear equations of the form of Equations 2a, b, there are analogies between the quantities describing these systems and quantities in the stress–strain relation. For example, strain behaves like voltage, stress like current, and the complex compliance $\hat{J}$ like the complex impedance of an electric circuit. Similar to the resonance phenomenon in circuits and mechanical systems, a Q can be defined by the average energy W per cycle divided by the energy lost or work done per cycle, ΔW:

$$Q = \frac{W}{\Delta W} \tag{4}$$

Large Q's imply small energy loss; small Q's imply large loss. Q is a measure of the area contained in the hysteresis loop of a stress–strain cycle. The inverse of (Equation 4), Q^{-1}, is sometimes simply termed the *attenuation or internal friction* (Knopoff, 1964*)*.

Plane waves of frequency ω and propagating in the + or − direction can be defined by the phasor $\exp(i\omega t - \hat{k}t)$ where $\hat{k}$ is a complex wave number $\frac{\omega}{\hat{c}}$ and $\hat{c}$ is a complex velocity defined from the local density ρ and complex modulus $\hat{G}$, with

$$\hat{c} = \sqrt{\frac{\hat{G}}{\rho}} \tag{5}$$

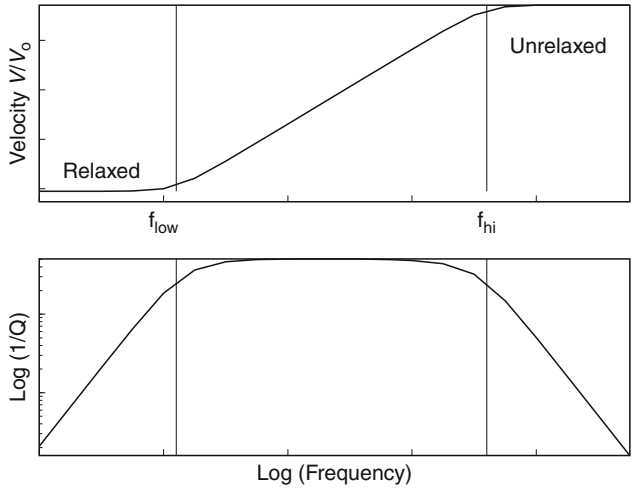
From the average energy density and loss per cycle of a complex plane wave, it can be shown that $Q = \frac{\mathrm{Re}(\hat{G})}{\mathrm{Im}(\hat{G})}$. It is often less confusing to report the reciprocal parameter Q^{-1}, which represents the usually small perturbations to perfect elasticity.

The $Q^{-1}(\omega)$ relaxation spectrum

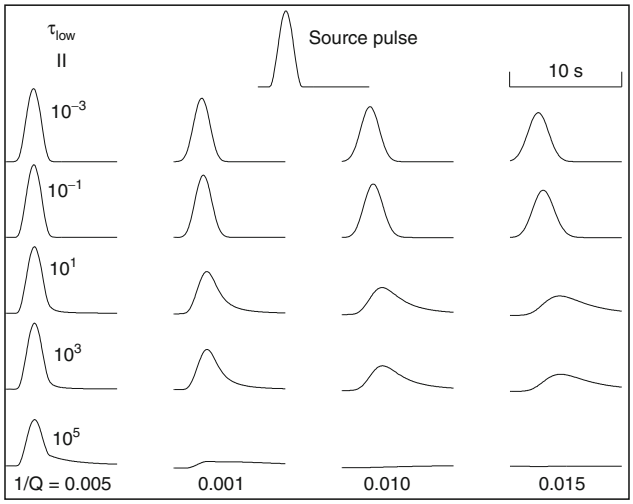
Since $\hat{G}$ depends on frequency, Q also depends on frequency. Zener (1960) described the frequency-dependent effects on an elastic modulus of a solid having a single characteristic time τ for the relaxation of stress. A distribution of relaxation times can be constructed to give a Q^{-1} having a general dependence on frequency. The function $Q^{-1}(\omega)$ is called the *relaxation spectrum*. In the Earth and in many solid materials, the relaxation spectrum is observed to be slowly varying and nearly constant over a broadband of frequencies. A theoretical requirement is that the attenuation Q^{-1} cannot increase faster than ω^1 or decrease faster than ω^{-1}. Figure 2 shows how a continuous distribution of relaxations can produce a Q^{-1} that is nearly constant with a frequency over a broadband. Once the limits of an absorption band are specified, however, it is not possible to have an arbitrarily high Q^{-1} (low viscoelastic Q) over an arbitrarily broad-frequency band without making an unrealistically large modulus defect ΔG. Measured modulus defects in shear are typically less than 25%.

Velocity dispersion

Although the dispersion in elastic moduli had long been known and predicted from the theories of viscoelasticity,



Seismic, Viscoelastic Attenuation, Figure 2 Viscoelastic dispersion of seismic velocity (*top*) and attenuation (*bottom*) showing a relaxation spectrum constant with frequency between two corner frequencies.



Seismic, Viscoelastic Attenuation, Figure 3 Pulse distortion showing the effects of viscoelastic dispersion for variable low-frequency corner and peak attenuation.

it was only widely recognized in seismology when velocity models determined in the low-frequency band from the normal modes of the earth (0.0001–0.01 Hz) were compared with velocity models determined in a high-frequency band (0.1–10 Hz) of body waves (Dziewonski and Anderson, 1981). The models were found to differ and the difference was found to agree with the amount of dispersion predicted from average Q models of the Earth. For example, since the preliminary reference Earth model (PREM), was derived from observations of both the travel times of body waves as well as the eigen frequencies of free oscillations, it reports velocities referenced at both 0.001 Hz and at 1 Hz.

Another more subtle effect of this velocity dispersion can be seen in the propagation of pulses as body waves. A stress disturbance that propagates from its point of initiation as a symmetric narrow Gaussian or triangle-shaped function in time gradually evolves into an asymmetric pulse (Figure 3). High frequencies traveling faster than low frequencies are preferentially loaded into the front of the pulse (Futterman, 1962; Carpenter, 1967). Common theories for the physical mechanism of earthquakes as either frictional slip on a plane or a propagating crack triggered by tectonic stress often predict a far-field displacement pulse that has either a different or opposite form of asymmetry than that predicted for the effect of viscoelastic attenuation. These differences can assist in separating the effects of the source-time history from the effects of viscoelastic attenuation.

Effects of scattering

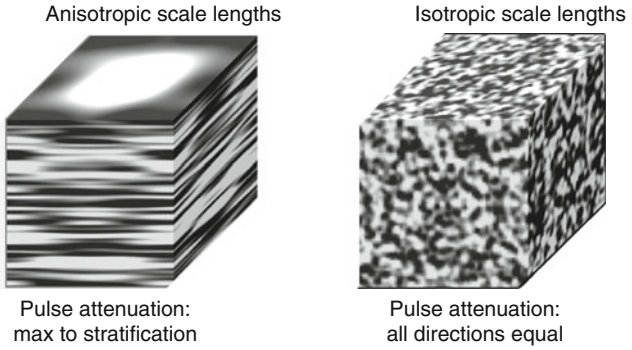
Equivalent medium

At frequencies that are so low that wavelengths are much larger than the characteristic scales of heterogeneity, the attenuative effects of scattering can usually be neglected. At sufficiently low frequency, little energy is lost to scattering, and the medium behaves like an equivalent medium, having properties that are an average of small-scale heterogeneities.

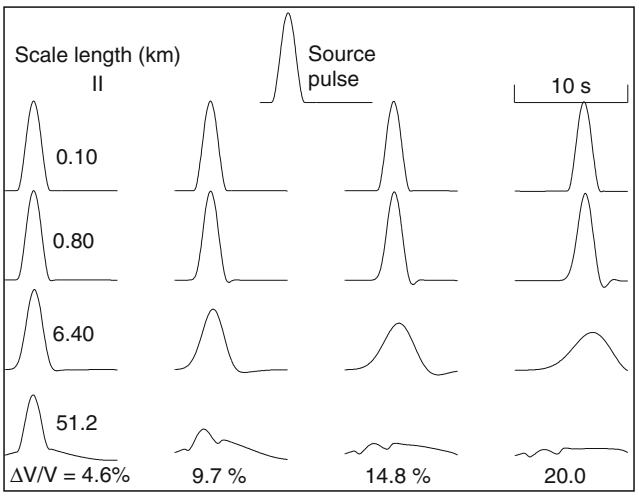
Stochastic dispersion

The most complicated domain in which to perform calculations is where the wavelength is of the order of the scale length of the heterogeneity (Figure 4). In this domain, the presence of heterogeneities can profoundly alter the propagation of the wavefield, both the initial cycle of a body wave pulse as well as the motion immediately following the initial cycle or *coda*. The effects of scattering can be calculated in a one-dimensional medium consisting of thin planar layers in which the velocity in each layer is assigned randomly (O'Doherty and Anstey, 1971; Richards and Menke, 1983). A prediction of such experiments is that body waves will exhibit a stochastic dispersion in which high-frequency energy is transferred into the coda following the first several cycles. This stochastic dispersion may have some biasing effects on measures of intrinsic attenuation. In measures of the spectrum taken over a narrow time window, different results can be obtained, depending on the length of window analyzed, with less attenuation of higher frequencies estimated from longer time windows.

Pulse measurements such as width and rise time may also be biased because higher-frequency energy has been transferred out of the pulse into the later coda. This behavior is opposite to the effects of intrinsic attenuation on a propagating pulse, in which higher frequencies arrive at the beginning of the pulse. A symmetrically shaped displacement source pulse loses less of its symmetry as it propagates through the heterogeneous medium (Figure 5). Anisotropy of the scale lengths of heterogeneity can also be important factor (Hong and Wu, 2005), attenuation being strongest for paths for which the



Seismic, Viscoelastic Attenuation, Figure 4 Example heterogeneity in the Earth and the directional dependence of attenuation of a body wave pulse for wavelengths that are approximately equal to either the dominant scale length (isotropic heterogeneity) or the dominant scale length in the direction of propagation (anisotropic heterogeneity).



Seismic, Viscoelastic Attenuation, Figure 5 Pulse distortion showing the effects of scattering attenuation for variable scale lengths and velocity perturbation calculated by Cormier and Li (2002) using the Dynamic Composite Elastic Modulus (DYCEM) theory of Kaelin and Johnson (1998).

wavelength is of the order of the characteristic scale length in the medium in that direction.

Effects of anisotropy

The existence of general anisotropy in the real part of the elastic modulus has the potential to bias some estimates of anelastic attenuation from either shear wave pulses or surface waves. In a medium having general anisotropy, the decompositions of shear wave motion into SH and SV motion will each contain the interference of two orthogonal shear wave polarizations that are neither SH nor SV (see *Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting*). The broadening of the SH component due to the interference of two quasi-S waves arriving close in time can be mistaken for the broadening due to anelastic attenuation. The regions of the deep Earth characterized by the strongest elastic anisotropy are the upper 400 km of the mantle (Silver, 1996) and the lowermost 400 km of the mantle near the core-mantle boundary (Panning and Romanowiz, 2006). The effects of elastic anisotropy must be removed by combined analysis of SV and SH components of motion, resolving the polarizations of two quasi-S waves, before viscoelastic attenuation can be properly measured.

Measurement and modeling attenuation

Measurements of amplitude of seismic waves may be taken directly from seismograms or from their frequency spectra. To measure the attenuation, we must predict its effects from a model and vary the parameters of the model to fit the observed amplitude, amplitude ratio, or waveform. The effects of intrinsic attenuation in any modeling algorithm operating in the frequency domain can be simply obtained by allowing elastic moduli and/propagation velocities to become complex. Elastic boundary conditions, reflection and transmission at boundaries, travel times, and amplitudes are calculated exactly as in a non-attenuating solid but with elastic moduli and associated velocities analytically continued to complex values. This step of analytic continuation of real moduli to complex moduli is the same whether one wishes to predict the waveform of a body wave or surface wave or spectrum of free oscillations. The size of the imaginary part of the elastic moduli, parameterized by the value of Q^{-1}as a function of depth and frequency, is chosen to match an observed waveform, spectrum, amplitude ratio, or spectral ratio.

The attenuation operator for body waves

As an example of these procedures, consider an experiment with body waves. The effects on a body wave of source radiation, geometric spreading, reflection-transmission, and intrinsic attenuation are most conveniently expressed in the frequency domain by a product of complex functions. The complex $\widehat{\Omega}(\vec{x},\omega)$ spectrum of a body wave propagation from a point $\vec{x}_o$ to a receiver at $\vec{x}$ is

$$\widehat{\Omega}(\vec{x},\omega) = \widehat{B}(\vec{x}_o,\vec{x},\omega)\,\widehat{S}(\omega)\,\widehat{A}(\omega) \tag{6}$$

The function $\widehat{S}(\omega)$ is the Fourier transform of the source-time function. $\widehat{B}(\vec{x}_o,\vec{x},\omega)$ incorporates a product of reflection-transmission coefficients, reverberations at source and receiver, geometric spreading, and source radiation pattern. $\widehat{A}(\omega)$ is defined by

$$\widehat{A}(\omega) = \exp[i\omega\,\widehat{T}(\omega)] \tag{7}$$

where $\widehat{T}(\omega)$ is the complex travel time obtained by integrating the reciprocal of complex velocity along a ray or normal to the wave front of the body wave:

$$\widehat{T}(\omega) = \int_{\text{ray}} \widehat{c}(\omega)\mathrm{d}s \tag{8}$$

For body waves, the dominant effect of attenuation on amplitude and phase is given by $\hat{A}(\omega)$. The effects of attenuation on reflection-transmission coefficients and geometric spreading, which have been lumped into $\hat{B}$ are much smaller and can be neglected unless the attenuation is very large (Q is very small). For $Q >> 1$, $\hat{A}(\omega)$ can be rewritten as

$$\hat{A}(\omega) = \exp\left[\frac{-\omega t*(\omega)}{2}\right] \exp\left\{i\omega\left[\mathrm{Re}\ \hat{T}(\infty) - \frac{H[t*(\omega)]}{2}\right]\right\} \tag{9}$$

where

$$t*(\omega) = \int_{\text{ray}} \frac{Q^{-1}}{\hat{c}(\omega)} ds \tag{10}$$

In Equation 9, the attenuation effect is contained in the factor $\exp\left[\frac{-\omega t*(\omega)}{2}\right]$, and the dispersive effect is in the factor $\exp\left\{i\omega\left[\mathrm{Re}\ \hat{T}(\infty) - \frac{H[t*(\omega)]}{2}\right]\right\}$. The operator H is a Hilbert transform. In a band of frequencies in which Q and t^* are nearly constant

$$H[t*(\omega)]/2 = \frac{\ln(\omega/\omega_0)}{\pi}\ t* \tag{11}$$

where ω_0 is a reference frequency contained in the band (Liu et al., 1976). The value of $T(\infty)$ need not be known and can be replaced by some reference time or predicted from an Earth model for the phase being analyzed. The Hilbert transform relation in Equation 11 for the dispersive phase of $\hat{A}(\omega)$ says that $\hat{A}(\omega)$ must be a minimum phase filter in the frequency domain. In general, the Fourier transform of the source-time function, $\hat{S}(\omega)$, is not a minimum phase filter, which can help in the separation and discrimination of the source spectrum from the effects of $\hat{A}(\omega)$ in the total expression for the far-field spectrum $\Omega(\vec{x}, \omega)$.

The phase given by Equation 11 will be accurate only between and far from the low- and high-frequency corners of the relaxation spectrum. Accurate representations of $\hat{A}(\omega)$ across a broad-frequency band can be obtained for general relaxation spectra by substituting expressions for complex velocity $\hat{c}(\omega)$ in Equation 8 obtained by superposing multiple Zener relaxations centered on single relaxation times whose strength is varied to achieve a desired shape for the relaxation spectrum. A useful expression for $\hat{c}(\omega)$ that is accurate for all frequencies across a relaxation spectrum, which is flat between two corner frequencies, can be derived from formulae for complex modulus given by Minster (1978), and is

$$\hat{c}(\omega) = c_{\text{ref}}(\omega_0)\frac{\sqrt{1 + 2\pi Q^{-1}\ln[\psi(\omega)]}}{\mathrm{Re}\sqrt{1 + 2\pi Q^{-1}\ln[\psi(\omega_0)]}} \tag{12a}$$

where

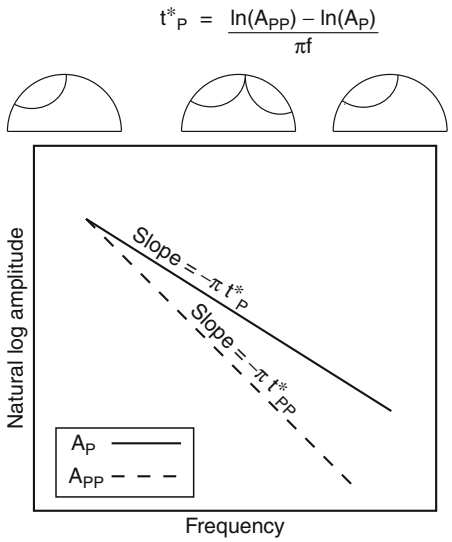
$$\psi(\omega) = \frac{i\omega + 1/\tau_1}{i\omega + 1/\tau_2} \tag{12b}$$

with τ_1 and τ_2 the relaxation times corresponding to the low and high frequency corners respectively. $c_{ref}(\omega_0)$ is a real velocity at the reference frequency ω_0.

Most measurements of attenuation attempt to measure only the amplitude effect of attenuation through the term $\exp[-\omega t*(\omega)/2]$ from the spectral shape of body waves. There are basically two types of experiments commonly reported: matching of (1) spectral decay rates and (2) spectral ratios. In experiment (1) a shape for the displacement source spectrum $S(\omega)$ is assumed usually to be a flat level followed by a decay of ω^{-2} above a corner frequency. The additional decay observed at high frequencies in data spectra is taken as a measure of $t*$ in $\exp[-\omega t*(\omega)/2]$. In experiment (2), a ratio of two different seismic phases from the same source is observed in which the source spectrum is assumed to approximately cancel and the factor related to ratios of geometric spreading and near source and receiver crustal reverberations can be assumed to contribute a simple constant scalar factor. If the phases analyzed are recorded at the same receiver and are incident at nearly the same angles, then crustal reverberations at the source and receiver will approximately cancel. Both types of experiments usually apply some type of smoothing to the spectra to remove biasing effects of spectral holes caused by interfering crustal multiples, source complexities, scattering, and multipathing that are not included in the simple propagation model. Figure 6 illustrates an attenuation experiment of this type.

Since t^* measures only the path-integrated effect of attenuation, many such experiments for different ray paths, bottoming at a range of different depths, are needed to construct a model of Q as a function of depth. Serious breakdowns in this approach, however, exist for cases in which the factorization of the observed spectrum into a product of a geometric spreading, source spectrum, and crustal effects is no longer accurate. One such case is when the body waves in question experience frequency-dependent effects of diffraction near caustics or grazing incidence to discontinuities. The spectral ratios of PKnKP waves, for example, are dominated by the effects of frequency-dependent reflection and transmission coefficients at grazing incidence to the core-mantle boundary. Instead of decreasing linearly with increasing frequency, an observed spectral ratio increases with frequency and exhibits a curvature in a log-log plot, which is consistent with a Q near infinity ($Q^{-1} = 0$) in the other core (Cormier and Richards, 1976).

It is becoming more common to model and invert for viscoelastic attenuation parameters in the time domain, including not only the magnitude of the viscoelastic attenuation parameter Q^{-1}, but also its frequency dependence. Examples of such a study are the inversions for Q^{-1} in the inner core assuming either a viscoelastic



Seismic, Viscoelastic Attenuation, Figure 6 The steps (*top*) to measure the path-integrated attenuation t^* of P waves in the mantle from a log-log plot (*bottom*) of stacked PP and P spectra (A_{PP} and A_P). The distances of observed P and PP spectra are chosen such that each turning ray path of PP is identical in shape and length to that of the single turning ray path of P in the mantle (Adapted from figures in Warren and Shearer, 2000).

(Li and Cormier, 2002) or a scattering origin of attenuation (Cormier and Li, 2002). In these studies, the combined effects of mantle attenuation and source-time function were first modeled by fitting P waves observed in the great circle range 30–90°. Attenuation in the liquid outer core was assumed to be zero. Parameters defining a viscoelastic relaxation spectrum in the inner core were then varied to match the observed PKIKP waveforms. Care must be taken to examine a broad range of attenuation parameters because waveform inversions of this type are very non-linear.

Free oscillations and surface waves

Measurements of attenuation in the low-frequency band of the free oscillations of the Earth are conducted in the frequency domain by observing the width of the individual resonance peaks associated with each mode. These measurements face special problems associated with the broadening produced by lateral heterogeneity of elastic Earth structure. This heterogeneity splits the degenerate modes of a radially symmetric Earth, making a set of modes that would have the same frequency have slightly different frequencies. The slightly different frequencies of the split modes may not be easily resolved in the data spectra and can be confused with the broadening of a single resonance peak of a mode caused by attenuation.

Lateral heterogeneity also complicates the measurement of viscoelastic attenuation of surface waves. Heterogeneity introduces focusing, defocusing, and multipathing, all of which must be accurately modeled to understand the separate attenuative effects of viscoelasticity.

The frequency band of free-oscillation and surface waves (0.001–0.1 Hz), however, offers the best hope of obtaining radially symmetric whole-Earth models of viscoelastic attenuation in this frequency band. This is because lateral variations in attenuation structure are averaged by the gravest modes of oscillation and surface waves that make multiple circuits around the Earth. Computational advances have made the division between free-oscillation and surface wave studies fuzzier, with common approaches now amounting to time-domain modeling of complete low-frequency (<0.1 Hz) seismograms for combined three-dimensional models of viscoelasticity and heterogeneity.

Numerical modeling

Fully numerical modeling of the seismic wavefield allows the combined effects of heterogeneity and viscoelasticity in three-dimensions to be predicted. If the numerical technique is formulated in the frequency domain, substituting a complex velocity for an assumed relaxation spectrum can incorporate viscoelastic attenuation.

If the technique is formulated in the time domain by a finite difference approach, it is neither simple nor efficient to incorporate attenuation by convolution of the wavefield calculated in a non-attenuating medium with an attenuation operator $A(t)$ for individual waves propagating in the attenuating medium, where $A(t)$ is the Fourier transform of $\widehat{A}(\omega)$ defined in Equation 9. Instead, time-domain *memory functions* can be defined to describe a viscoelastic relaxation (Robertsson et al., 1994; Blanch et al., 1995) that can be integrated over time simultaneously with the equations describing particle velocity or displacement and stress. In practice, only three-memory functions, distributed evenly over the logarithm of their characteristic times, are required to simulate a broad frequency band in which Q^{-1} varies slowly.

Interpretation of attenuation measurements in the earth

Shear versus bulk attenuation

In the most general theory of viscoleasticity, it is possible to have with energy loss to occur during both a cycle of volumetric strain as well as shear strain. Since the velocity of a P wave depends on both the bulk and shear moduli, the attenuation Q_P^{-1} of a P wave can be written as a linear combination of the attenuations Q_K^{-1} and Q_S^{-1} defined from complex shear and bulk moduli:

$$Q_P^{-1} = L\,Q_S^{-1} + (1 - L)\,Q_K^{-1} \tag{13}$$

where $(4/3)(V_S/V_P)^2$ and V_P and V_S are the compressional and shear velocities respectively (Anderson, 1989). Although plausible mechanisms for defects in bulk moduli have been found in both laboratory measurements and analytic models of specific attenuation mechanisms, measurements on real data find that bulk dissipation in

the earth is small and, in most cases, can be neglected. One exception may occur when the pressure and temperature state in a narrow depth regions of the earth are close to those near a phase transition, either solid–liquid (Stevenson, 1983) or solid–solid (Ricard and Chambat, 2009). Except for these regions, intrinsic attenuation occurs almost entirely in shear, associated with lateral movement of lattice defects, grain boundaries and/or fluids rather than with changes in material volume. Hence, for viscoelastic attenuation purely in shear in a Poisson solid, for which $V_P = \sqrt{3}\, V_S$,

$$Q_P^{-1} = \frac{4}{9}\, Q_s^{-1} \tag{14}$$

and the parameter for path-integrated attenuation of S waves or $t*_S$ is approximately $4\, t*_P$. Most experiments confirm these values. There is a suggestion, however, that the apparent Q_P^{-1} tends to approach Q_S^{-1} and $t*_S < 4\, t*_P$ at frequencies higher than 1 Hz. These observations are likely evidence of scattering rather than of bulk attenuation because the effects of scattering increase at higher frequencies. With scattering, the apparent Q_S^{-1} tends to approach the apparent Q_P^{-1}, especially when they are measured from pulse widths or spectra taken in a frequency band and medium for which wavelengths are of the order of richest heterogeneity scale lengths of the medium. Thus, the assumption of viscoelastic attenuation occurring mainly in shear can aid in separating the effects of scattering from intrinsic attenuation in body wave pulses.

Frequency dependence

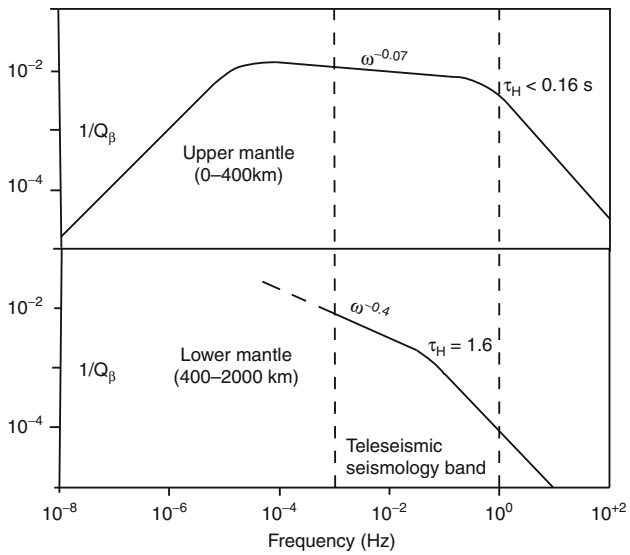
When the results of attenuation measurements determined from free-oscillations and body waves in the 0.0001–0.1 Hz band began to be compared with observations of body wave spectra in the 1–10 Hz band, it became apparent that even under the assumption of a white source spectrum that an increase in Q with frequency was necessary to explain the amplitude of spectra in the 1–10 Hz band.

Thermal activation

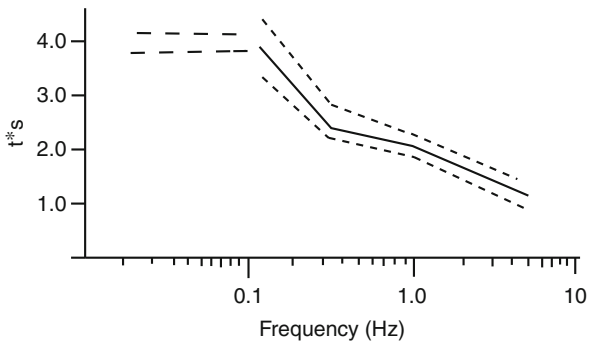
Frequency dependence of viscoelastic attenuation has been interpreted in terms of physical mechanisms of attenuation that are thermally activated. In these mechanisms, the low-frequency corner f_L is tied to a relaxation time τ_L, where $f_L = 1/(2\,\pi\,\tau_L)$. The time τ_L depends on temperature T and pressure P as follows:

$$\tau_L = \tau_0 \exp\left(\frac{E* + PV*}{RT}\right) \tag{15}$$

where $E*$ and $V*$ are the activation energy and volume, respectively. Both the low- and high-frequency corners (f_L, f_H) of an absorption band are assumed to be similarly affected, temperature and pressure acting to slide the absorption band through a band of frequencies. A typical width to expect for the relaxation spectrum of the mantle is about 5 orders of magnitude in frequency, $\tau_L/\tau_H = 10^5$ (Minster and Anderson, 1981; Anderson and Given, 1982). A simplified model of an absorption band with depth in the earth's mantle is shown in Figure 7. The movement of the absorption band toward lower frequencies longer periods in the mantle below 400 km depth is consistent with the type of behavior shown in Figure 8 for the $t*_S$ measured from shear waves of an earthquake. The difference in the location of the absorption band with respect to the band of seismic frequencies is consistent with models of the temperature and pressure profiles of



Seismic, Viscoelastic Attenuation, Figure 7 A frequency and depth-dependent model of shear attenuation in the Earth's mantle derived from modeling broadband shear waves. This model has been used by the National Earthquake Information Center (NEIC) to correct for viscoelastic attenuation in the reported radiated elastic energy from earthquakes (Boatwright and Choy, 1986; Choy and Boatwright, 1995).



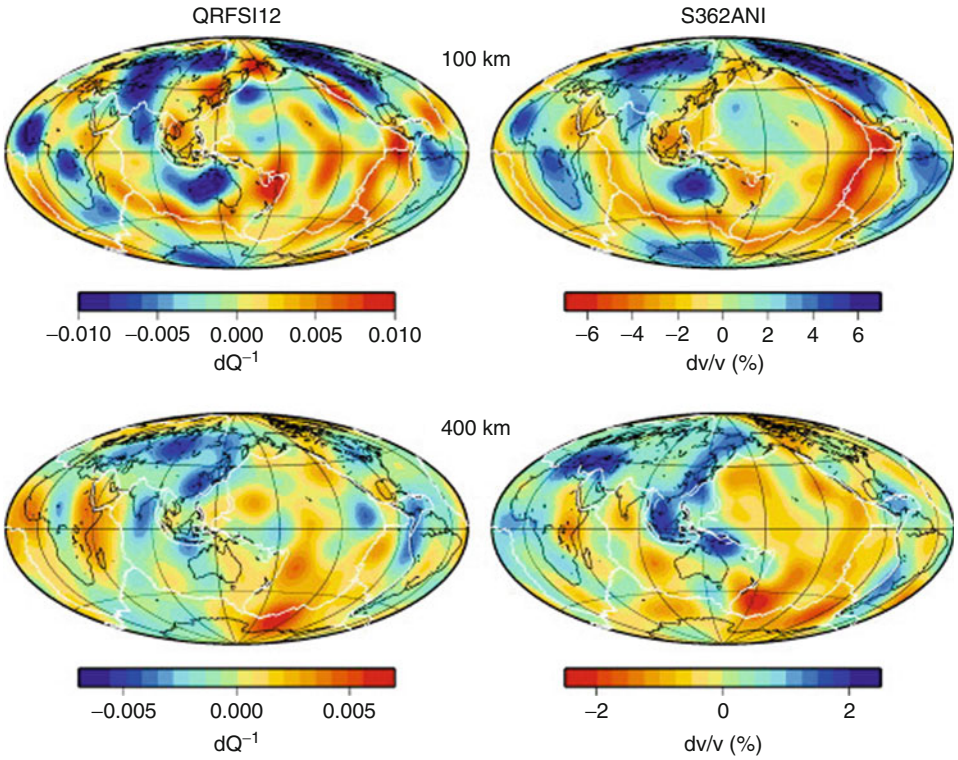
Seismic, Viscoelastic Attenuation, Figure 8 Path-integrated attenuation $t*$ of S waves in the mantle as a function of frequency determined from modeling broadband shear waves predicted from the frequency and depth-dependent attenuation model shown in Figure 7 (Choy and Cormier, 1986).

the Earth's mantle for specific values of E^* and V^*. Relaxation times are also affected by the grain size of minerals, which may increase from millimeter to centimeter in the upper 400 km of the mantle (Faul and Jackson, 2005). A rapid increase in temperature with depth can rapidly change the location of the absorption band with respect to the seismic band. Given a specific temperature profile, estimated values of the activation energy and volume, and grain size, a combined shear velocity and Q_S profile can be predicted and modified to fit an observed shear velocity profile.

Regional variations

Romanowicz and Mitchell (2007) review and interpret both global and many regional variations in intrinsic attenuation, including correlations with velocity perturbations. Tomographic images of perturbations to seismic velocities and attenuations in the mantle can qualitatively be interpreted as images of lateral temperature variations, leaving open the possibility of additional contributions to the observed heterogeneity from chemical variations. In the upper mantle, tectonically active regions overlain by radiogenically younger crust are more attenuating than the mantle underlying inactive regions such as continental shields (Figure 9). The shape of the frequency dependence across the seismic band seems to remain similar in different regions, although the Q at a given frequency is lower for a tectonically young region than for an older shield region. Although motion of interstitial water and partial melt can produce high attenuation, grain boundary and defect mechanisms in a dry mantle have been shown to be equally effective in explaining regions of high attenuation in the upper mantle, except possibly behind island arcs and directly beneath active spreading ridges (Karato and Jung, 1998; Faul and Jackson, 2005).

Generally, perturbations in attenuation Q^{-1} inversely correlate with those in shear wave velocity (Roth et al., 2000). The correlations between shear velocity and shear attenuation appear to be consistent with thermal activation, in which the dispersive effect of attenuation acts jointly with variations in the high-frequency corner of the mantle relaxation spectrum to produce the observed variations in travel time and frequency content. Deep chemical differences between the upper mantle beneath shields and that beneath young continents and oceans as well as in the deep mantle, however, have been suggested by comparing anomalies in shear velocity versus bulk velocity V_K, where $V_K = \sqrt{V_P^2 - (4/3)V_S^2}$. Milder lateral temperature differences in the mid- and lower mantle tend to make the relaxation spectrum more laterally stable in height, width, and location within a frequency band, reducing the observed lateral heterogeneity in velocity and attenuation in these regions.



Seismic, Viscoelastic Attenuation, Figure 9 Inverted upper mantle shear velocity perturbations and shear attenuation. (From Dalton et al., 2009).

Global models of attenuation (e.g., Gung and Romanowicz, 2004) often do not have the resolving power to detect spatially concentrated regions of high attenuation and sharp spatial gradients found in regional studies beneath and near island arcs, mid-ocean spreading ridges, and hot mantle plumes. The dense path coverage required of higher-frequency (0.1–2 Hz) body waves to resolve smaller spatial scales usually is lacking, except in regions containing dense seismic networks. Even larger scale, long-established lateral variations, such as high attenuation west of the Rocky Mountains in North America and low-attenuation east (Der et al., 1982), are not very apparent in some global studies (Warren and Shearer, 2002).

Strain dependence

Laboratory measurements of Q in rocks find dependence in strain beginning at strains of about 10^{-6}. The strain dependence decreases with increasing confining pressure. The Q is also strongly dependent on moisture and interstitial fluids between cracks in rocks and grain boundaries and soils. These observations are consistent with a physical mechanism of frictional sliding across cracks. Unlike viscoelastic relaxations, which are representative of all linear mechanisms, frictional sliding is an inherently non-linear mechanism, depending on strain amplitude.

Estimates, when non-linear effects occur, may be made by calculating the strain associated with the seismic wave being analyzed. A rough estimate can be obtained by assuming that the wave front is a plane wave and dividing the particle velocity by the propagation velocity. For example, the particle velocities of body waves observed in strong ground motion recordings from 0 km to 10 km from the hypocenter of a magnitude 6 earthquake are typically 0.01 m/s. If the body wave propagates at 3 km/s, the strain observed at the strong ground motion site is roughly 0.01 m/s divided by 3×10^3 m/s or strain = 3.3×10^{-5}. This value is likely to be in the non-linear regime of surficial rocks having open cracks or pores. In this strain regime, it becomes important to solve the elastic equation of motion with non-linear terms in its rheology (Bonilla et al., 2005), including terms proportional to the square of strain.

Summary

The intrinsic attenuation of seismic waves in the earth has been found to be consistent with loss mechanisms that are thermally activated. The observed regional and frequency dependences of seismic Q agree with the expected lateral variations in a geotherm having a rapid temperature increase in the upper 400 km of the mantle, followed by a slower vertical and lateral variation in the mid- and lower mantle. High velocities correlate with regions of low attenuation; low seismic velocities correlate with regions of high attenuation. Measurements are consistent with losses primarily in shear rather than bulk deformations.

The existence of lateral heterogeneity in the elastic properties of the Earth complicates the measurement of viscoelastic properties. The longer scale lengths of heterogeneity can split modes of free oscillation and focus and defocus body waves and surface waves. Shorter scale lengths scatter seismic energy, broaden the waveforms of body waves, and redistribute energy into different time and angular windows. Observations that are useful for discriminating between the effects of scattering attenuation versus viscoelastic attenuation include the ratio of apparent P wave attenuation to apparent S wave attenuation, the rate of velocity dispersion within a frequency band, and the apparent viscoelastic modulus defect. The intensity of heterogeneity in percent fluctuation of velocities and densities is higher at shorter scale lengths at shallower depths in the Earth's crust and upper mantle. There is still a need for experiments that determine finer details of how the distribution of heterogeneity changes with depth and lateral location in the Earth and its anisotropy of scale lengths. Many, if not most experiments, have not completely removed the effects of heterogeneity on the apparent attenuation, making their results an upper bound on the viscoelastic Q^{-1}.

Laboratory experiments find a transition from linear to non-linear rheology at strains of the order of 10^{-6}. The observed strain dependence of Q and its dependence on pressure in the shallow crust agree with a mechanism of frictional sliding of cracks. It is still unknown how and at what strain levels linear superposition begins to break down close to a seismic source.

Although a consensus has been reached on the major features and thermal activation of intrinsic attenuation in most of the Earth's upper mantle, this is less true of other deep regions of the Earth. Definitive experiments are still needed for the determination of Q^{-1} in the lowermost 400 km of the mantle, where increased lateral heterogeneity exists across a broad spatial spectrum, complicating the separation of its effects from those of viscoelasticity. A concept unifying lateral variations in velocity, elastic anisotropy, scattering, and apparent attenuation in the uppermost inner core is needed (e.g., Calvet and Margerin, 2008).

Bibliography

Anderson, D. L., 1989. *Theory of the Earth*. Boston: Blackwell Scientific Publications.

Anderson, D. L., and Given, J. W., 1982. The absorption band Q model for the Earth. *Journal of Geophysical Research*, **87**, 3893–3904.

Blanch, J. O., Robertsson, J. O. A., and Symes, W. W., 1995. Optimally efficient constant Q modeling. *Geophysics*, **60**, 176–184.

Boatwright, J., and Choy, G., 1986. Teleseismic estimates of the energy radiated by shallow earthquakes. *Journal of Geophysical Research*, **91**, 2095–2112.

Bonilla, L. F., Archuleta, R. J., and Lavallée, D., 2005. Hysteretic and dilatant behavior of cohesionless soils and their effects on nonlinear site response: field data, observations and modeling. *Bulletin. Seismological Society of America*, **95**, 2373–2395.

Calvet, M., and Margerin, L., 2008. Constraints on grain size and stable iron phases in the uppermost inner core from multiple scattering modeling of seismic velocity and attenuation. *Earth and Planetary Science Letters*, **267**, 200–212.

Carpenter, E. W., 1967. Teleseismic signal calculated for underground, underwater, and atmospheric explosions. *Geophysics*, **32**, 17–32.
Choy, G. L., and Boatwright, J. L., 1995. Global patterns of radiated seismic energy and apparent stress. *Journal of Geophysical Research*, **100**, 18205–18228.
Choy, G. L., and Cormier, V. F., 1986. Direct measurement of the mantle attenuation operator from broadband P and S waves. *Journal of Geophysical Research*, **91**, 7326–7342.
Cormier, V. F., and Li, X., 2002. Frequency dependent attenuation in the inner core: Part II. A scattering and fabric interpretation. *Journal of Geophysical Research,* **107**(B12), doi:10.1029/2002JB1796.
Cormier, V. F., and Richards, P. G., 1976. Comments on "The Damping of Core Waves" by Anthony Qamar and Alfredo Eisenberg. *Journal of Geophysical Research*, **81**, 3066–3068.
Dalton, C. A., Ekstrom, G., and Dziewonski, A. M., 2009. Global seismological shear velocity and attenuation: A comparison with experimental observations. *Earth and Planetary Science Letters*, **284**, 65–75.
Der, Z. A., McElfresh, T. W., and O'Dannell, 1982. An investigation of regional variations and frequency dependence of anelastic attenuation in the United States in the 0.5–4 Hz. band. *Geophysical Journal of the Royal Astronomical Society*, **69**, 67–100.
Dziewonski, A. M., and Anderson, D. L., 1981. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, **24**, 297–356.
Faul, U. H., and Jackson, I., 2005. The seismological signature of temperature and grain size variations in the upper mantle. *Earth and Planetary Science Letters*, **234**, 119–134.
Futterman, W. I., 1962. Dispersive body waves. *Journal of Geophysical Research*, **67**, 5279–5291.
Gross, B., 1953. *Mathematical Structure of the Theories of Viscoelasticity*. Paris: Hermann.
Gung, Y., and Romanowiz, B. A., 2004. Q tomography of the upper mantle using three component long period waveforms. *Geophysical Journal International*, **147**, 831–830.
Hong, T.-K., and Wu, R.-S., 2005. Scattering of elastic waves in geometrically anisotropic random media and its implication to sounding of heterogeneity in the Earth's deep interior. *Geophysical Journal International*, **163**, 324–338.
Jackson, I., 1993. Progress in the experimental study of seismic attenuation. *Annual Review of Earth and Planetary Sciences*, **21**, 375–406.
Jackson, I., 2007. Properties of rocks and minerals – physical origin of anelasticity and attenuation in rocks. In: Schubert, G. (ed.), *Treatise on Geophysics*, 2, Amsterdam: Elsevier, pp. 493–525.
Jackson, D. D., and Anderson, D. L., 1970. Physical mechanisms of seismic wave attenuation. *Reviews of Geophysics and Space Physics*, **8**, 1–63.
Jackson, I., Webb, S., Weston, L., and Boness, D., 2005. Frequency dependence of elastic wave speeds at high temperature: a direct experimental demonstration. *Physics of the Earth and Planetary Interiors*, **148**, 85–96.
Kaelin, B., and Johnson, L. R., 1998. Dynamic composite elastic medium theory. Part II. Three-dimensional media. *Journal of Applied Physics*, **84**, 5458–5468.
Karato, S.-I., and Jung, H., 1998. Water partial melting and the origin of the seismic low velocity zone in the upper mantle. *Earth and Planetary Science Letters*, **157**, 193–207.
Knopoff, L., 1964. Q. *Review of Geophysics*, **2**(4), 625–660.
Kohistedt, D. L., 2007. Properties of rocks and minerals – constitutive equations, rheological behavior, and viscosity of rocks. In: Schubert, G. (ed.), *Treatise on Geophysics*, 2, Amsterdam: Elsevier, 390–417.
Li, X., and Cormier, V. F., 2002. Frequency dependent attenuation in the inner core: Part I. A viscoelastic interpretation. *Journal of Geophysical Research*, **107**(B12), doi:10.1029/2002JB001795.
Liu, H-P., Anderson, D. L., and Kanamori, H., 1976. Velocity dispersion due to anelasticity: implications for seismology and mantle composition. *Geophysical Journal of the Royal Astronomical Society*, **47**, 41–58.
Minster, J. B., 1978. Transient and impulse responses of a one-dimensional linearly attenuating medium—I. Analytical results. *Geophysical Journal of the Royal Astronomical Society*, **52**, 479–501.
Minster, B., and Anderson, D. L., 1981. A model of dislocation-controlled rheology for the mantle. *Philosophical Transactions of the Royal Society of London*, **299**, 319–356.
Nowick, A. S., and Berry, B. S., 1972. *Anelastic Relaxation in Crystalline Solids*. New York: Academic, p. 677.
O'Connell, R. J., and Budiansky, B., 1977. Viscoelastic properties of fluid-saturated cracked solids. *Journal of Geophysical Research*, **82**, 5719–5735.
O'Doherty, R. F., and Anstey, N. A., 1971. Reflections on amplitudes. *Geophysical Prospecting*, **19**, 430–458.
Panning, M. P., and Romanowicz, B. A., 2006. A three dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophysical Journal International*, **167**, 361–379.
Ricard, Y., and Chambat, F., 2009. Seismic attenuation in a phase change coexistence loop. *Physics of the Earth and Planetary Interiors*, **176**, 124–131.
Richards, P. G., and Menke, W., 1983. The apparent attenuation of a scattering medium. *Bulletin. Seismological Society of America*, **73**, 1005–1021.
Robertsson, J. O. A., Blanch, J. O., and Symes, W. W., 1994. Viscoelastic finite-difference modeling. *Geophysics*, **59**, 1444–1456.
Romanowicz, B., and Mitchell, B., 2007. Deep earth structure: Q of the earth from crust to core. In: Schubert, G. (ed.), *Treatise on Geophysics*, 1, Amsterdam: Elsevier, 731–774.
Roth, E. G., Wiens, D. A., and Zhao, D., 2000. An empirical relationship betweeen seismic attenuation and velocity anomalies in the upper mantle. *Geophysical Research Letters*, **27**, 601–604.
Silver, P. G., 1996. Seismic anisotropy beneath the continents: Probing the depths of geology. *Annual Review of Earth and Planetary Sciences*, **24**, 385–432.
Stevenson, D. J., 1983. Anomalous bulk viscosity of two-phase fluids and implications for planetary interiors. *Journal of Geophysical Research*, **88**, 2445–2455.
Warren, L. M., and Shearer, P. M., 2000. Investigating the frequency dependence of mantle Q by stacking P and PP spectra. *Journal of Geophysical Research*, **105**(B11), 25391–25402.
Warren, L. M., and Shearer, P. M., 2002. Mapping lateral variations in upper mantle attenuation by stacking P and PP spectra. *Journal of Geophysical Research*, **107**(B12), 2342.
Zener, C., 1960. *Elasticity and Anelasticity of Metals*. Chicago: The University of Chicago Press.

Cross-references

Body Waves
Earth's Structure, Core
Earth's Structure, Lower Mantle
Earth's Structure, Upper Mantle
Earthquakes, Early and Strong Motion Warning
Earthquakes, Strong-Ground Motion
Energy Partitioning of Seismic Waves
Free Oscillations of the Earth
Lithosphere, Mechanical Properties
Mantle D″ Layer
Mantle Viscosity
Numerical Methods, Finite Difference

Propagation of Elastic Waves: Fundamentals
Seismic Anisotropy
Seismic Diffraction
Seismic Monitoring of Nuclear Explosions
Seismic Phase Names: IASPEI Standard
Seismic Properties of Rocks
Seismic Tomography
Seismic Velocity-Temperature Relationships
Seismic Wave Propagation in Real Media: Numerical Modeling Approaches
Seismic Waves, Scattering
Seismic, Ray Theory
Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting
Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat

SEISMIC, WAVEFORM MODELING AND TOMOGRAPHY

Yanghua Wang
Department of Earth Science and Engineering, Centre for Reservoir Geophysics, Imperial College London, London, UK

Synonyms

Waveform inversion

Definition

Tomography. An inversion method to produce slicing image of the internal structures of an object, by recording wavefield propagating through and scattered/refracted/reflected back from the object, and observing the difference in the effects on the wave energy impinging on those structures.

Introduction

Waveform tomography is an imaging technology using seismic data to reconstruct the Earth subsurface physical properties, represented by seismic velocity, attenuation coefficient, etc. It is usually formulated as an inverse problem. Many seismic tomography methods use the arrival time information extracted from the waveform recordings as the input data to an inversion process. These methods are generically referred to as travel time tomography. Comparing to travel time tomography, waveform tomography has at least two major features. First, the input data consist of the seismic waveforms themselves, as opposed to travel times, amplitudes, or some other secondary attributes of the recorded data. This makes waveform tomography have a better resolution than travel time tomography. Secondly, the underlying numerical method is based on the full wave equation, as opposed to a ray approximation or a Born approximation. This makes waveform tomography more accurate than the travel time counterpart. These two features also make the inverse problem of waveform tomography more difficult to solve.

However, as field seismic data usually lack low-frequency information, travel time tomography may provide a reliable starting model for the iterative waveform inversion (Pratt et al., 2002). If waveform tomography is implemented in the frequency domain, the inversion process proceeds from low to high-frequency components. In the time domain, a band-pass filter can be applied to the seismic data, and inversion uses band-pass filtered data with low frequencies first and proceeds to higher frequencies. Even for the frequency-domain implementation, in order to suppress the low signal-to-noise ratio effect of frequency data and to make the inversion procedure robust, waveform tomography usually uses a group of frequencies simultaneously in an iterative inversion, and proceeds from low to high-frequency groups, generating a high-resolution image of the subsurface model.

Waveform tomography has been used successfully on transmission data, such as crosshole seismic data, and to wide-angle reflection/refraction seismic data with a certain degree of success. It has also been used for regional-scale studies on the crustal and upper mantle velocity structure, using scattering waves, surface waves and SH-waves from either exploration seismics or broad-band teleseismograms. In addition, there is a good progress on its application to reflection-seismic data with limited source-receiver offsets (Wang and Rao, 2009). The seismic reflection method is a routine practice in the hydrocarbon exploration, and the data dominated by the pre-critical reflection energy, reflected back from subsurface contrasts in physical parameters, are well suited for seismic migration for the structural image. Applying the waveform tomography technique to these reflection data can quantitatively extract the geophysical parameters, for identifying different lithologies and different fracture characteristics and even for indicating the hydrocarbon distribution directly.

Waveform modeling

Seismic wave propagation satisfies both Newton's second law and Hook's law, and can be expressed by the following five simultaneous first-order differential equations in two-dimensional case:

$$
\begin{aligned}
\rho \frac{\partial u}{\partial t} &= \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xz}}{\partial z}, \\
\rho \frac{\partial w}{\partial t} &= \frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{zz}}{\partial z}, \\
\frac{\partial \tau_{xx}}{\partial t} &= c_{11} \frac{\partial u}{\partial x} + c_{13} \frac{\partial w}{\partial z} + c_{15} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right), \\
\frac{\partial \tau_{zz}}{\partial t} &= c_{31} \frac{\partial u}{\partial x} + c_{33} \frac{\partial w}{\partial z} + c_{35} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right), \\
\frac{\partial \tau_{xz}}{\partial t} &= c_{51} \frac{\partial u}{\partial x} + c_{53} \frac{\partial w}{\partial z} + c_{55} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right),
\end{aligned}
\tag{1}
$$

where u and w are the particle-velocity components in the horizontal and vertical directions, respectively, τ_{xx}, τ_{xz} and τ_{zz} are stress components, ρ is density, and

$$\begin{bmatrix} c_{11} & c_{13} & c_{15} \\ c_{31} & c_{33} & c_{35} \\ c_{51} & c_{53} & c_{55} \end{bmatrix} \tag{2}$$

are the elastic constants in Hook's law relating stress to strain in the two-dimensional case. The first two equations are derived from Newton's second law, and the last three from Hooke's law for an elastic medium. Eliminating stress components, these five simultaneous equations can be coupled into two equations,

$$\begin{aligned} &-\rho\frac{\partial^2 u}{\partial t^2}+\frac{\partial}{\partial x}\left(c_{11}\frac{\partial u}{\partial x}+c_{15}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial x}\left(c_{15}\frac{\partial u}{\partial z}+c_{13}\frac{\partial w}{\partial z}\right)\\ &+\frac{\partial}{\partial z}\left(c_{51}\frac{\partial u}{\partial x}+c_{55}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial z}\left(c_{55}\frac{\partial u}{\partial z}+c_{53}\frac{\partial w}{\partial z}\right)=0,\\ &-\rho\frac{\partial^2 w}{\partial t^2}+\frac{\partial}{\partial x}\left(c_{51}\frac{\partial u}{\partial x}+c_{55}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial x}\left(c_{55}\frac{\partial u}{\partial z}+c_{53}\frac{\partial w}{\partial z}\right)\\ &+\frac{\partial}{\partial z}\left(c_{31}\frac{\partial u}{\partial x}+c_{35}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial z}\left(c_{35}\frac{\partial u}{\partial z}+c_{33}\frac{\partial w}{\partial z}\right)=0, \end{aligned} \tag{3}$$

defined in terms of horizontal and vertical wave components u and w. These equations are valid for the Earth media with arbitrary anisotropy and heterogeneity.

In the forward modeling of seismic wave propagation based upon either five or two simultaneous equations, there has been steady development in the finite-difference implementations. As the time derivatives are approximated by $\partial u/\partial t \approx (u_{i+1}-u_i)/\Delta t$ and $\partial^2 u/\partial t^2 \approx (u_{i+1}-2u_i+u_{i-1})/\Delta t^2$, where i is the time index, $t_i = i\Delta t$, and Δt is the step size in time, the wavefield at the current time can be calculated based on the previous wavefield: $u_{i+1} = f(u_i)$ or $u_{i+1} = f(u_i, u_{i-1})$; it is therefore an explicit finite-difference scheme (Alford et al., 1974; Kelly et al., 1976). For the spatial derivatives, a second-order staggered grid scheme for numerical stability was proposed by Virieux (1986), and was extended to a fourth-order scheme for greater accuracy by Levander (1988).

In the frequency domain, the two simultaneous equations become

$$\begin{aligned} &\omega^2\rho u+\frac{\partial}{\partial x}\left(c_{11}\frac{\partial u}{\partial x}+c_{15}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial x}\left(c_{15}\frac{\partial u}{\partial z}+c_{13}\frac{\partial w}{\partial z}\right)\\ &+\frac{\partial}{\partial z}\left(c_{51}\frac{\partial u}{\partial x}+c_{55}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial z}\left(c_{55}\frac{\partial u}{\partial z}+c_{53}\frac{\partial w}{\partial z}\right)=0,\\ &\omega^2\rho w+\frac{\partial}{\partial x}\left(c_{51}\frac{\partial u}{\partial x}+c_{55}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial x}\left(c_{55}\frac{\partial u}{\partial z}+c_{53}\frac{\partial w}{\partial z}\right)\\ &+\frac{\partial}{\partial z}\left(c_{31}\frac{\partial u}{\partial x}+c_{35}\frac{\partial w}{\partial x}\right)+\frac{\partial}{\partial z}\left(c_{35}\frac{\partial u}{\partial z}+c_{33}\frac{\partial w}{\partial z}\right)=0. \end{aligned} \tag{4}$$

where ω is the angular frequency. This system of wave equations can be solved in parallel for individual frequencies, using also a finite-difference method. Once all frequency components of the wavefield at any spatial position are obtained, performing inverse Fourier transform with respect to the frequency variable will produce the time-domain seismic trace at this location.

However, the frequency-domain implementation is much more time consuming, compared to its counterpart in the time domain. This is because the frequency-domain calculation involves the solution of linear algebraic equations as the follows. With a finite-differencing scheme, Equation 4 may be presented in a matrix form as

$$\mathbf{Mu} = \mathbf{s}, \tag{5}$$

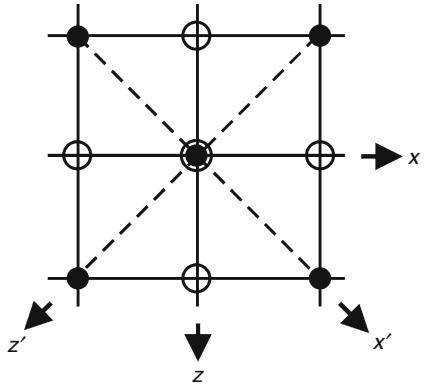
where $\mathbf{M}$ is a matrix approximating the partial differential operators, $\mathbf{u}$ is a vector representing the two components of the wavefield at all grids, and $\mathbf{s}$ is a vector representing the source term, which is zero everywhere except at the location of the source (added to the right-hand side of the wave equations). To obtain wavefield $\mathbf{u}$, one needs to solve the linear algebraic equation system 5, that is, to solve the inverse matrix, $\mathbf{M}^{-1}$. It is therefore an implicit finite-difference scheme.

The matrix $\mathbf{M}$ represents a significant storage requirement, which is largely determined by the numerical bandwidth of $\mathbf{M}$ and by the manner in which the structural sparsity of the matrix is maintained in any solution method. The bandwidth of the differencing matrix is determined by the number of nodes needed for a spatial derivative. For example, for a second-order finite-differencing, a first derivative needs 3×3 nodes, and for a fourth-order 5×5 nodes. A minimal, rotated computational star is a scheme that can minimize the bandwidth of the matrix and meanwhile improve the accuracy for finite-difference modeling. It does not require any new grid points as used in a second-order finite-differencing but can produce an equivalent of fourth-order accuracy in finite-differencing (Štekl and Pratt, 1998). As shown in Figure 1, the ○ symbol represents the five nodes needed in second-order finite-differencing of a first derivative, and the • symbol represents the five nodes in a 45° rotated finite-differencing star. Coupling these two second-order finite-differencing stars does not use additional points outside the 3×3 nodes, and there will be no increase in the numerical bandwidth of the differencing matrix. Therefore, the increment in computational cost and storage requirement over the ordinary second-order scheme is negligible.

In a rotated coordinate system, the elastic constants (2) become

$$\begin{bmatrix} c_{11}' & c_{13}' & c_{15}' \\ c_{31}' & c_{33}' & c_{35}' \\ c_{51}' & c_{53}' & c_{55}' \end{bmatrix} = \mathbf{B}\begin{bmatrix} c_{11} & c_{13} & c_{15} \\ c_{31} & c_{33} & c_{35} \\ c_{51} & c_{53} & c_{55} \end{bmatrix}\mathbf{B}^T, \tag{6}$$

where the rotation operator is



Seismic, Waveform Modeling and Tomography, Figure 1 The minimal, rotated computational star for finite-difference modeling. The symbol ○ represents the only five nodes required in the ordinary, second-order finite-differencing star. The symbol • represents the five nodes in a rotated computational star. The combination of these two stars indicates the coupling of the central node to the nearest neighbors on the grid. The final star does not use additional points outside the 3×3 nodes, but coupling includes those nodes not presented in the original star.

$$\mathbf{B} = \begin{bmatrix} \cos^2\theta & \sin^2\theta & 2\sin\theta\cos\theta \\ \sin^2\theta & \cos^2\theta & 2\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix}, \tag{7}$$

and θ is rotated angle of new coordinate. Considering a special case of 45° rotation, where the rotated coordinate system is (x', z') and displacements are (u', w') in the rotated coordinate system, Equation 4 becomes

$$\begin{aligned} &\omega^2\rho u' + \frac{\partial}{\partial x'}\left(c_{11}'\frac{\partial u'}{\partial x'} + c_{15}'\frac{\partial w'}{\partial x'}\right) + \frac{\partial}{\partial x'}\left(c_{15}'\frac{\partial u'}{\partial z'} + c_{13}'\frac{\partial w'}{\partial z'}\right) \\ &+ \frac{\partial}{\partial z'}\left(c_{51}'\frac{\partial u'}{\partial x'} + c_{55}'\frac{\partial w'}{\partial x'}\right) + \frac{\partial}{\partial z'}\left(c_{55}'\frac{\partial u'}{\partial z'} + c_{53}'\frac{\partial w'}{\partial z'}\right) = 0, \\ &\omega^2\rho w' + \frac{\partial}{\partial x'}\left(c_{51}'\frac{\partial u'}{\partial x'} + c_{55}'\frac{\partial w'}{\partial x'}\right) + \frac{\partial}{\partial x'}\left(c_{55}'\frac{\partial u'}{\partial z'} + c_{53}'\frac{\partial w'}{\partial z'}\right) \\ &+ \frac{\partial}{\partial z'}\left(c_{31}'\frac{\partial u'}{\partial x'} + c_{35}'\frac{\partial w'}{\partial x'}\right) + \frac{\partial}{\partial z'}\left(c_{35}'\frac{\partial u'}{\partial z'} + c_{33}'\frac{\partial w'}{\partial z'}\right) = 0. \end{aligned} \tag{8}$$

The relationship between displacements (u', w') in the rotated coordinate system and (u, w) in the original coordinate system is given by

$$u = \frac{1}{\sqrt{2}}(u' - w'),\ w = \frac{1}{\sqrt{2}}(u' + w'). \tag{9}$$

By subtracting and adding two equations in Equation 8 and then using relations in Equation 9, one can obtain

$$\begin{aligned} &\omega^2\rho u + \frac{1}{2}\frac{\partial}{\partial x'}\left((c_{11}' - c_{51}' - c_{15}' + c_{55}')\frac{\partial u}{\partial x'}\right. \\ &\quad\left. + (c_{11}' - c_{51}' + c_{15}' - c_{55}')\frac{\partial w}{\partial x'}\right) \\ &\quad + \frac{1}{2}\frac{\partial}{\partial x'}\left((c_{15}' - c_{55}' - c_{13}' + c_{53}')\frac{\partial u}{\partial z'}\right. \\ &\quad\left. + (c_{15}' - c_{55}' + c_{13}' - c_{53}')\frac{\partial w}{\partial z'}\right) \\ &\quad + \frac{1}{2}\frac{\partial}{\partial z'}\left((c_{51}' - c_{31}' - c_{55}' + c_{35}')\frac{\partial u}{\partial x'}\right. \\ &\quad\left. + (c_{51}' - c_{31}' + c_{55}' - c_{35}')\frac{\partial w}{\partial x'}\right) \\ &\quad + \frac{1}{2}\frac{\partial}{\partial z'}\left((c_{55}' - c_{35}' - c_{53}' + c_{33}')\frac{\partial u}{\partial z'}\right. \\ &\quad\left. + (c_{55}' - c_{35}' + c_{53}' - c_{33}')\frac{\partial w}{\partial z'}\right) = 0, \\ &\omega^2\rho w + \frac{1}{2}\frac{\partial}{\partial x'}\left((c_{11}' + c_{51}' - c_{15}' - c_{55}')\frac{\partial u}{\partial x'}\right. \\ &\quad\left. + (c_{11}' + c_{51}' + c_{15}' + c_{55}')\frac{\partial w}{\partial x'}\right) \\ &\quad + \frac{1}{2}\frac{\partial}{\partial x'}\left((c_{15}' + c_{55}' - c_{13}' - c_{53}')\frac{\partial u}{\partial z'}\right. \\ &\quad\left. + (c_{15}' + c_{55}' + c_{13}' + c_{53}')\frac{\partial w}{\partial z'}\right) \\ &\quad + \frac{1}{2}\frac{\partial}{\partial z'}\left((c_{51}' + c_{31}' - c_{55}' - c_{35}')\frac{\partial u}{\partial x'}\right. \\ &\quad\left. + (c_{51}' + c_{31}' + c_{55}' + c_{35}')\frac{\partial w}{\partial x'}\right) \\ &\quad + \frac{1}{2}\frac{\partial}{\partial z'}\left((c_{55}' + c_{35}' - c_{53}' - c_{33}')\frac{\partial u}{\partial z'}\right. \\ &\quad\left. + (c_{55}' + c_{35}' + c_{53}' + c_{33}')\frac{\partial w}{\partial z'}\right) = 0. \end{aligned} \tag{10}$$

Now there are two systems of partial differential equations:

$$\begin{aligned} \rho\omega^2 u + A_1 &= 0, \\ \rho\omega^2 w + B_1 &= 0, \end{aligned} \quad \text{and} \quad \begin{aligned} \rho\omega^2 u + A_2 &= 0, \\ \rho\omega^2 w + B_2 &= 0, \end{aligned} \tag{11}$$

where A_1 and B_1 are the partial differential parts of Equation 4 in the original coordinate system, and A_2 and B_2 are the partial differential parts of equation system 10 in the rotated coordinate system. The two systems can be combined to as

$$\begin{aligned} \rho\omega^2 u + aA_1 + (1-a)A_2 &= 0, \\ \rho\omega^2 w + aB_1 + (1-a)B_2 &= 0, \end{aligned} \tag{12}$$

where a is an optimal real-valued coefficient that must be sought to maximize the accuracy of the solution for all propagation directions.

While it is feasible to calculate the inverse matrix $\mathbf{M}^{-1}$ for a small, 2D model, in many practical problems such as 3D seismic modeling, the system $\mathbf{Mu} = \mathbf{s}$ with a large, sparse matrix is solved iteratively rather than directly. An approximate solution $\tilde{\mathbf{u}}$ is first obtained by solving using a related, preconditioning matrix. This preconditioner is chosen such that it is related to the true matrix, but is much faster to invert. The approximate solution $\tilde{\mathbf{u}}$ from the preconditioner is then substituted into the true matrix equation to discover the effective source, $\tilde{\mathbf{s}} = \mathbf{M}\tilde{\mathbf{u}}$. The difference $\mathbf{s} - \tilde{\mathbf{s}}$ between this effective source and the true source can then be treated as an actual source for the next iteration. The final solution $\mathbf{u}$ to the original matrix equation is found if the difference is sufficiently small (Warner et al., 2008).

Figure 2 shows the snapshot of wave propagation through an anisotropic medium. The elastic coefficients are summarized as

$$c_{11} = 29.26, \quad c_{33} = 23.98, \quad c_{55} = 4.29,$$
$$c_{13} = c_{31} = 17.27, \quad c_{15} = c_{51} = 1.98,$$
$$c_{35} = c_{53} = 2.53(\text{Gpa}),$$

and the density of the medium is a constant of 2.5 g/cm^3. The source is a Ricker wavelet with dominate frequency of 10 Hz. The snapshot shows clearly qP-wave front propagating in difference directions with different velocities. S-wave front can also been observed in these figures.

Waveform modeling uses an elastic wave equation, for the generality, and produces a wavefield presented as particle-velocity components. In a homogeneous, isotropic media, $c_{11} = c_{33} = \lambda + 2\mu$, $c_{13} = c_{31} = \lambda$ and $c_{55} = \mu$, where λ and μ are Lamé parameters, and $c_{15} = c_{51} = c_{35} = c_{53} = 0$. In an acoustic assumption, $c_{11} = c_{33} = c_{13} = \lambda$ and $c_{55} = c_{15} = c_{35} = 0$. With the acoustic assumption, most field seismic records are pressure data. According to Newton's second law, the relationship between particle velocity (u, w) and pressure P is

$$\rho \frac{\partial u}{\partial t} = -\frac{\partial P}{\partial x}, \rho \frac{\partial w}{\partial t} = -\frac{\partial P}{\partial z}. \tag{13}$$

Inverse method

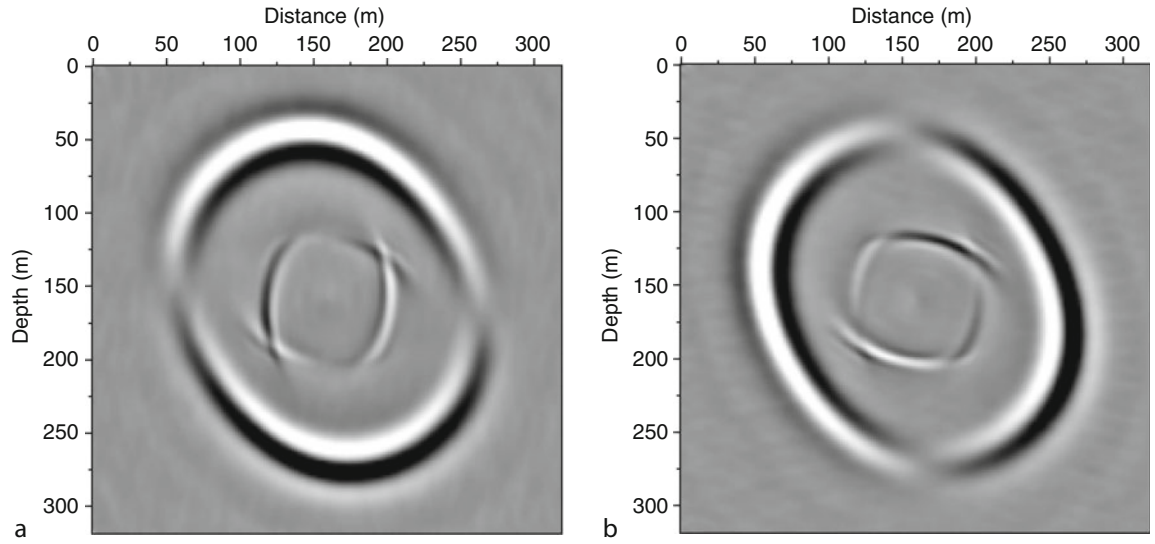
Seismic waveform tomography is an inverse problem with an objective function defined by

$$\varphi(\mathbf{m}) = [\mathbf{P}(\mathbf{m}) - \mathbf{P}_{\text{obs}}]^{\text{H}} \mathbf{C}_{\text{D}}^{-1} [\mathbf{P}(\mathbf{m}) - \mathbf{P}_{\text{obs}}] + \mu[\mathbf{m} - \mathbf{m}_0]^{\text{H}} \mathbf{C}_{\text{M}}^{-1} [\mathbf{m} - \mathbf{m}_0], \tag{14}$$

where $\mathbf{m}_0$ is a reference model, $\mathbf{m}$ is the model to invert for, $\mathbf{P}(\mathbf{m})$ is a modeled data set based on model $\mathbf{m}$, $\mathbf{P}_{\text{obs}}$ is an observed data set, $\mathbf{C}_{\text{D}}$ is the data covariance matrix with units of $(data)^2$, defining the uncertainties in the data set, $\mathbf{C}_{\text{M}}$ is the model covariance matrix with units of $(model\ parameter)^2$, and μ is a scalar that controls the relative weights of the data fitting term and the model constraint in the objective function. In Equation 14, the superscript $^{\text{H}}$ denotes the complex conjugate transpose.

For minimizing the objective function, a gradient method can be used. It starts with the differentiation of the objective function with respect to the model parameters:

$$\frac{\partial \varphi}{\partial \mathbf{m}} = 2(\mathbf{F}^{\text{H}} \mathbf{C}_{\text{D}}^{-1} \delta\mathbf{P} + \mu \mathbf{C}_{\text{M}}^{-1} \delta\mathbf{m}), \tag{15}$$



Seismic, Waveform Modeling and Tomography, Figure 2 Snap shot of wave propagation in an anisotropic medium: (**a**) horizontal component and (**b**) vertical component.

where $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ is the model perturbation, $\delta\mathbf{P} = \mathbf{P}(\mathbf{m}) - \mathbf{P}_{obs}$ is the data residual, and $\mathbf{F}$ is a matrix of the Fréchet derivative of $\mathbf{P}(\mathbf{m})$ with respect to the model $\mathbf{m}$. The first term in Equation 15 is the gradient direction of the data misfit:

$$\hat{\boldsymbol{\gamma}} = \mathbf{F}^{\mathrm{H}}\mathbf{C}_{\mathrm{D}}^{-1}\delta\mathbf{P} = \mathbf{F}^{\mathrm{H}}\delta\hat{\mathbf{P}}, \tag{16}$$

where $\delta\hat{\mathbf{P}} = \mathbf{C}_{\mathrm{D}}^{-1}\delta\mathbf{P}$ is a weighted data residual. Set $\partial\varphi/\partial\mathbf{m} = 0$ in Equation 15, one obtains the following equation

$$\delta\mathbf{m} = -\alpha\mathbf{C}_{\mathrm{M}}\hat{\boldsymbol{\gamma}}, \tag{17}$$

where α is a update step length that needs to be determined.

In order to evaluate the gradient $\hat{\boldsymbol{\gamma}}$ using Equation 16, one needs to know the Fréchet matrix $\mathbf{F}$, which is obtained from the following linear formula,

$$\delta\mathbf{P} = \mathbf{F}\delta\mathbf{m}. \tag{18}$$

This is the first term in a Taylor's series for $\delta\mathbf{P}$ and relates the data perturbation $\delta\mathbf{P}$ to the model perturbation $\delta\mathbf{m}$. However, a direct computation of $[\mathbf{F}]_{ij} = \partial P_i/\partial m_j$ is a formidable task when P_i are seismic waveforms. The action of matrix $\mathbf{F}^{\mathrm{H}}$ on the weighted data residual vector $\delta\hat{\mathbf{P}}$ (Equation 16) can be computed by a series of forward modeling steps, summarized as follows (Lailly, 1984; Tarantola, 1984, 1987).

The frequency-domain acoustic wave equation for a constant density medium with velocity $c_0(\mathbf{r})$ is

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2(\mathbf{r})}\right)P_0(\mathbf{r},\omega) = -S(\omega)\delta(\mathbf{r}-\mathbf{r}_0), \tag{19}$$

where $\mathbf{r}$ is the position vector, $\mathbf{r}_0$ locates the source position, $S(\omega)$ is the source signature of frequency ω, and $P_0(\mathbf{r},\omega)$ is the (pressure) wavefield of this frequency and we drop off the frequency reference in the following discussion. If the velocity is perturbed by a small amount, $\delta c(\mathbf{r}) << c_0(\mathbf{r})$, that is, $c_0(\mathbf{r}) \to c(\mathbf{r}) = c_0(\mathbf{r}) + \delta c(\mathbf{r})$, then the total wavefield is correspondingly perturbed to $P_0(\mathbf{r}) \to P(\mathbf{r}) = P_0(\mathbf{r}) + \delta P(\mathbf{r})$. Following wave Equation 19, δP approximately satisfies

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2(\mathbf{r})}\right)\delta P(\mathbf{r}) = 2\omega^2 P_0(\mathbf{r})\frac{\delta c(\mathbf{r})}{c_0^3(\mathbf{r})}. \tag{20}$$

Considering $2\omega^2 P_0(\mathbf{r})\delta c(\mathbf{r})/c_0^3(\mathbf{r})$, the term on the right-hand side, as a series of "virtual sources" over $\mathbf{r}$, the integral solution for $\delta P(\mathbf{r})$ can be expressed as

$$\delta P(\mathbf{r}) = -\int_{\Omega} \delta c(\mathbf{r}')\frac{2\omega^2}{c_0^3(\mathbf{r}')}P_0(\mathbf{r}')G(\mathbf{r},\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{21}$$

where $G(\mathbf{r},\mathbf{r}')$ is the Green's function for the response at $\mathbf{r}$ to a point source at $\mathbf{r}'$ for the original velocity field. Note that in the acoustic case where one assumes density to be constant, and defines the model by the velocity field only, $\mathbf{m} \equiv \mathbf{c}$. Then comparing Equation 21 against the matrix-vector form of Equation 18, we see that the Fréchet matrix is defined with element, $F(\mathbf{r},\mathbf{r}') = -[2\omega^2/c_0^3(\mathbf{r}')]P_0(\mathbf{r}')G(\mathbf{r},\mathbf{r}')$. Substituting this Fréchet kernel into Equation 16, one obtains

$$\hat{\boldsymbol{\gamma}}(\mathbf{r}) = \left(\frac{2\omega^2}{c_0^3(\mathbf{r})}\right)^* \int_D P_0^*(\mathbf{r}')G^*(\mathbf{r},\mathbf{r}')\delta\hat{P}(\mathbf{r}'). \tag{22}$$

Replacing the integral over the data space with a summation over source and receiver pairs, denoted by s and g respectively, as the source and receiver position are inherently discrete and finite in number, one can obtain

$$\begin{aligned}\hat{\boldsymbol{\gamma}}(\mathbf{r}) &= \left(\frac{2\omega^2}{c_0^3(\mathbf{r})}\right)^* \sum_s \left(P_0^*(\mathbf{r};\mathbf{r}_s)\sum_g G^*(\mathbf{r},\mathbf{r}_g)\delta\hat{P}(\mathbf{r}_g;\mathbf{r}_s)\right)\\ &= \left(\frac{2\omega^2}{c_0^3(\mathbf{r})}\right)^* \sum_s \left(P_0^*(\mathbf{r};\mathbf{r}_s)P_b^*(\mathbf{r};\mathbf{r}_s)\right),\end{aligned} \tag{23}$$

where

$$P_b(\mathbf{r};\mathbf{r}_s) = \sum_g G(\mathbf{r},\mathbf{r}_g)\delta\hat{P}^*(\mathbf{r}_g;\mathbf{r}_s) \tag{24}$$

representing the wavefield generated by a series of virtual sources $\delta\hat{P}^*(\mathbf{r}_g)$, corresponding to a single source $\mathbf{r}_s$. Note that wavefield $P_b(\mathbf{r};\mathbf{r}_s)$ is not calculated directly from Equation 24, but is computed using the same forward modeling scheme as used for the wave Equation 19 with the virtual sources $\delta\hat{P}^*(\mathbf{r}_g)$, a procedure often referred to as data residual back-propagation.

In summary, waveform tomography is performed iteratively. For each iteration, the inversion procedure may be divided into four steps:

1. For a given model estimate, calculating the synthetic wavefield $P_0(\mathbf{r};\mathbf{r}_s)$ at space position $\mathbf{r}$ corresponding to a source point at $\mathbf{r}_s$
2. Using the weighted data residual $\delta\hat{\mathbf{P}} = \mathbf{C}_{\mathrm{D}}^{-1}\delta\mathbf{P}$ as virtual sources to generate a so-called back-propagation wavefield $P_{\mathrm{b}}(\mathbf{r};\mathbf{r}_s)$
3. Crosscorrelation of the original wavefield $P_0(\mathbf{r};\mathbf{r}_s)$ and the back-propagation wavefield $P_{\mathrm{b}}(\mathbf{r};\mathbf{r}_s)$ to get the gradient direction $\boldsymbol{\gamma} = \mathbf{C}_{\mathrm{M}}\hat{\boldsymbol{\gamma}}$, where $\mathbf{C}_{\mathrm{M}}$ is the model covariance matrix with units of $(\textit{model parameter})^2$
4. Estimating the model update $\delta\mathbf{m} = -\alpha\boldsymbol{\gamma}$, where α is the optimal step length that can be found by using the linear approximation or simply line search for a minimum of the objective function.

In a time-domain implementation of waveform inversion, the gradient direction is calculated by

$$\hat{\boldsymbol{\gamma}}(\mathbf{r}) = \frac{2}{c_0^3(\mathbf{r})}\sum_s \int \left(\frac{\partial}{\partial t}p_0(\mathbf{r};\mathbf{r}_s)\frac{\partial}{\partial t}p_b(\mathbf{r};\mathbf{r}_s)\right)\mathrm{d}t, \tag{25}$$

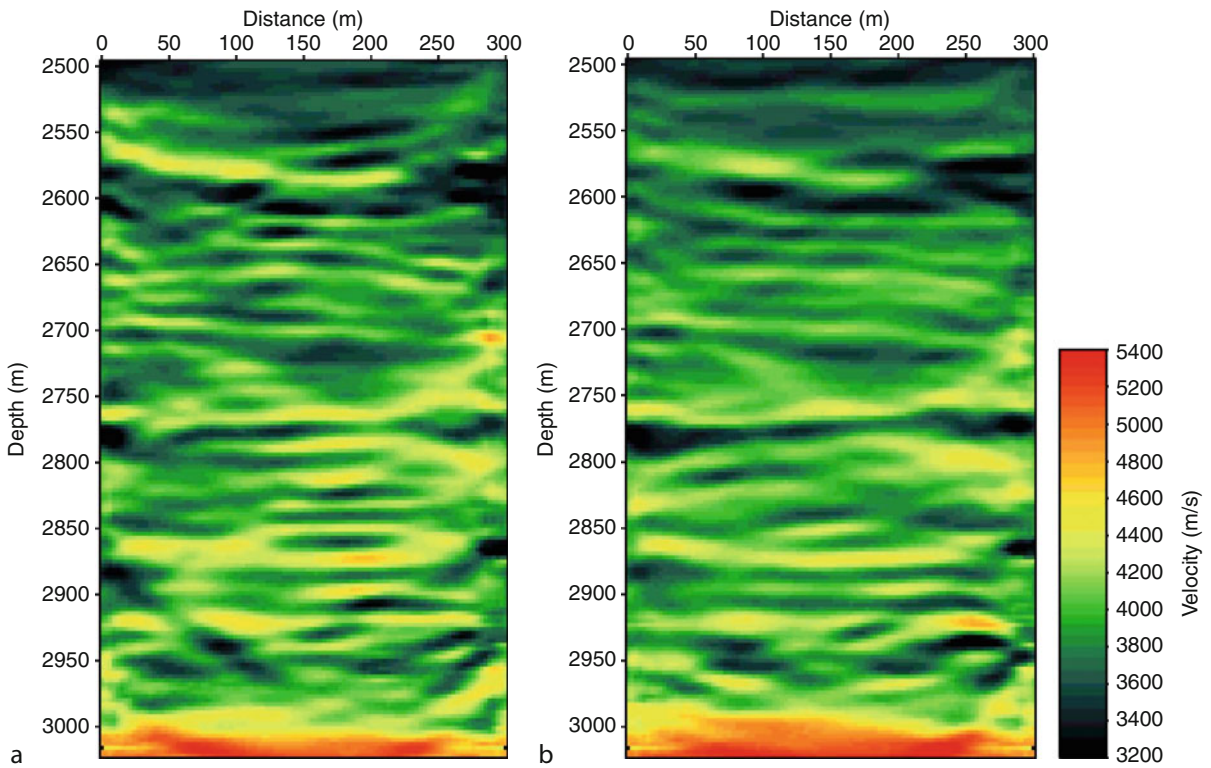
where $\partial p_0(\mathbf{r};\mathbf{r}_s)/\partial t$ denotes the time derivative of the time-domain forward propagated wavefields, and $\partial p_b(\mathbf{r};\mathbf{r}_s)/\partial t$ is the time derivative of the back-projected residual waveforms in the time domain. Zero-leg correlation of these two wavefields provides us the gradient direction for model updating.

Strategies for choosing frequencies

When dealing with real seismic data, poor signal-to-noise ratio of data slices in the frequency domain is a problem at least that affects waveform tomography. For real data application, a group of frequencies is necessarily used simultaneously for each individual iteration in the inversion procedure (Pratt and Shipp, 1999; Wang and Rao, 2006). Simultaneously using neighboring frequencies from the same spatial imaging position may have an averaging effect that suppresses the data noise to the input of the inversion. For a fixed number of model parameters to invert for, using many more data samples in the inversion means that the inverse problem becomes much better determined.

Here is a real data example of crosshole geometry, with two experiments to combat the noise in real data. In the first experiment, we use all selected frequencies consecutively (190, 195, 200, 205, . . . , 485 Hz). We start with the initial model generated from travel time inversion, and invert the 190 Hz data component first. Then, we switch to a higher frequency component (195 Hz) of the data as the inversion progresses. The result from each lower frequency is used as the starting model for the next higher frequency inversion. At each frequency stage, three iterations are carried out. Figure 3a shows the reconstructed image after using all 60 selected frequencies between 190 and 480 Hz with 5 Hz interval.

In the second experiment, we use a group of five neighboring frequencies simultaneously in the inversion. The 60 selected frequencies are assigned into 12 groups with increasing frequency contents. The result from each lower frequency group is used as the starting model for the inversion of the next higher frequency group. This strategy might mitigate the nonlinearity of the problem: For lower frequencies, the method is more tolerant of velocity errors, as these are less likely to lead to errors of more than a half-cycle in the waveforms (Pratt and Shipp, 1999). For each group, three iterations are carried out, proceeding through all groups. For each iteration, the gradient of each frequency group is computed using all five frequencies simultaneously. Figure 3b shows the tomographic image after using all of 12 frequency groups consecutively.



Seismic, Waveform Modeling and Tomography, Figure 3 Waveform tomography with a single frequency at a time or a group of frequencies simultaneously. (**a**) The inversion is executed by each frequency consecutively. The velocity image after using all 60 selected frequencies between 190 and 480 Hz (with 5 Hz interval). The image has strong X-shaped artifacts. (**b**) The inversion is executed one group by one group in sequence. The image after using all 12 frequency groups, each has five frequency components.

Comparing the images of those two experiments, one can observe that Figure 3a is marked by the presence of some X-shaped artifacts that cross the image. Such artifacts are quite often obtained in crosshole tomography, especially when waveform inversion is attempted. It is due to the non-uniform coverage of the object spectrum and the lack of information about the object spectrum in certain directions (Wu and Toksöz, 1987). When using multiple frequencies simultaneously, the inherent filtering (smoothing) effect might have an extrapolation effect of the object spectrum to the blind area. The second experiment has much fewer artifacts, and the image is smoother and more continuous than that of experiment one, especially at the 2,800–2,950 m portions. We recommend using the strategy of the second experiment in practice so that we can also mitigate the data noise effectively in the input of waveform tomography.

The computation time of the frequency-domain waveform tomography is linearly proportional to the total number of temporal frequencies used in the inversion. In reflection geometry, source-receiver pairs vary with different offsets and move along the surface. One may explore this coverage of a variety of plane-wave imaging directions, to reduce the number of frequencies needed in waveform tomography (Sirgue and Pratt, 2004). The frequency selection strategy is

$$f_{n+1} = \frac{f_n}{\cos\phi}, \tag{26}$$

where f_n is the frequency previously used, f_{n+1} is a new frequency for inversion, and $\cos\phi = z/\sqrt{h_{\max}^2 + z^2}$ is the cosine of the reflection angle, corresponding to the maximum value of the half source-receiver offset $h_{\max}$ and the target depth z. This strategy eliminates some frequencies but has a continuous coverage on vertical wave number k_z, as

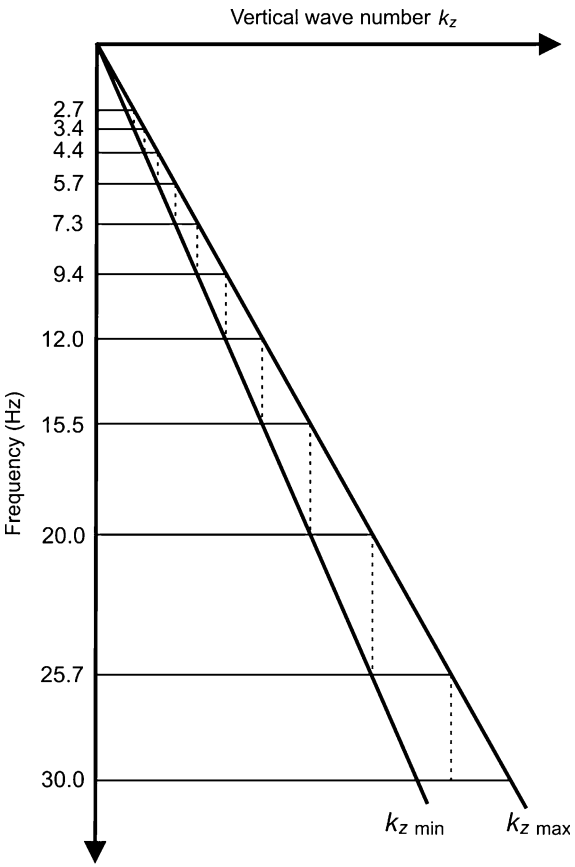
$$k_{z\min}(f_{n+1}) = k_{z\max}(f_n), \tag{27}$$

where

$$k_{z\max} = \frac{2f}{c_0}, \quad k_{z\min} = \frac{2f}{c_0}\cos\phi, \tag{28}$$

and $c_0/2$ are half of the background velocity. Figure 4 illustrates the wave number coverage corresponding to the selected frequencies in a waveform tomography example (Wang and Rao, 2009). For a model with $h_{\max} = 2,275\,\text{m}$, $z = 2,800\,\text{m}$ and $\cos\phi \approx 0.776$, the following 11 frequencies, 2.7, 3.4, 4.4, 5.7, 7.3, 9.4, 12.0, 15.5, 20.0, 25.7, and 30.0 Hz, may be selected. In this list, the dominant frequency 20 Hz is included in the inversion, and the last frequency is $f_{\max} = 30\,\text{Hz}$, due to the limitation of finite-difference modeling, which is less than the predicted value of 33.1 Hz.

On the other hand, if assuming there is no source-receiver offset coverage to be exploited, adequate wave number samples are needed in order to obtain a sufficiently good image in the model space. In this case,



Seismic, Waveform Modeling and Tomography, Figure 4 The coverage of the vertical wave numbers corresponding to 11 selected frequencies for an example waveform tomography.

the sampling rate should satisfy the anti-aliasing condition, $\Delta k_z \le 1/z_{\max}$, where $z_{\max}$ is the maximum depth to be imaged, and Δk_z is the sampling rate of the vertical wave number. Given the minimum value of the vertical wave number at frequency f by Equation 28, the difference between two neighboring wave number samples is

$$\Delta k_z \equiv k_{z\min}(f + \Delta f) - k_{z\min}(f) = 2\cos\phi\,\frac{\Delta f}{c_0} \le \frac{1}{z_{\max}}. \tag{29}$$

Finally, the following anti-aliasing condition for frequency sampling is obtained:

$$\Delta f \le \frac{c_0}{2z_{\max}\cos\phi}. \tag{30}$$

In this example, $\Delta f = 0.3\,\text{Hz}$.

Waveform tomography proceeds sequentially from low to high frequencies (Sirgue and Pratt, 2004; Pratt, 2008). This is because the nonlinearity of the inverse problem depends on the frequency of the data. As the misfit function at low frequencies is more linear than at high frequencies, the low-frequency inversion will have a better chance to

be successful and can accurately recover the low wave number components of the velocity model. The velocity model with accurate low wave number components is a good initial model for higher frequency inversions. At low frequencies, coarser grids can be used for computing numerical solutions of the wave equation than at high frequencies, resulting in a computational efficiency. This is a so-called multiscale approach (Bunks et al., 1995).

The multiscale approach can also be implemented in the time-domain waveform tomography (Boonyasiriwat et al., 2009). In the frequency domain, it is straightforward to apply the multiscale method, as a single frequency component is used at a time in the inversion. In the time-domain inversion, a frequency band is used instead of single frequency. For each band-passed filtered data set, the finite-difference grid size is determined by the maximum frequency of the band. Assuming a square grid, $\Delta x = \Delta z$, the grid size allowed to use in the forward modeling is determined by

$$\Delta x \leq \frac{\lambda_{\min}}{N} = \frac{c_{\min}}{N f_{\max}}, \tag{31}$$

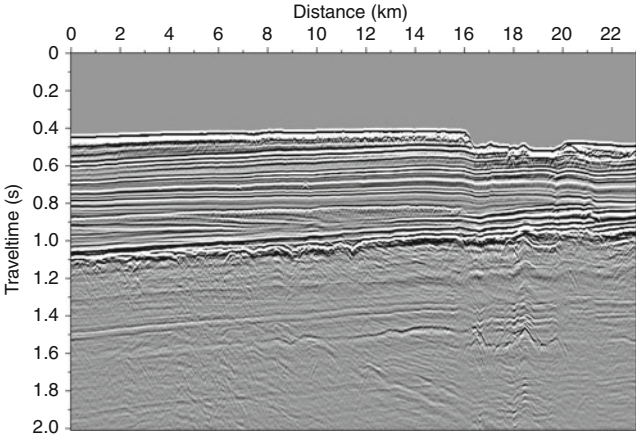
where $\lambda_{\min}$ is the minimum wavelength, $c_{\min}$ is the minimum velocity, $f_{\max}$ is the maximum frequency of the band, and N is the number of grid points at least per minimum wavelength, required by the numerical dispersion condition for the finite-difference scheme (Levander, 1988). At low frequencies, coarser grids can be used than at high frequencies. Therefore, low-frequency inversions will be considerably fast and efficient compared to high-frequency inversions, and can afford to take a large number of iterations in order to obtain an accurate estimate of low wave number components of the velocity model. Higher wave number components are progressively recovered through the sequential uses of higher frequency-band data in the inversions. Therefore, the multiscale approach should have a steady convergence than a single-scale method that tends to recover both low and high wave numbers simultaneously.

Preprocessing for field data waveform tomography

Let us demonstrate the application of waveform tomography to a real, marine seismic data set, as shown in Figure 5. This example profile shows the structure along the section. However, the input to waveform tomography is a series of shot records, which should be subject to some necessary processing, such as multiple attenuation and transformation of the original point sources to equivalent line sources, before they are used in tomography.

Figure 6a is a sample shot record, consisting of 120 traces with a minimum source-receiver offset of 337.5 m and a maximum source-receiver offset of 1,825 m. We investigate the feasibility of reflection-seismic tomography within such a narrow source-receiver offset range.

Waveform tomography usually does not include free-surface multiples. Otherwise including seismic multiples



Seismic, Waveform Modeling and Tomography, Figure 5 An example profile of marine seismic data, which shows the structure along a section. However, the input to waveform tomography is a series of shot records, which should be subject to some necessary processing.

bouncing back and forward within the water layer in the tomographic inversion will increase the nonlinearity of the problem. As the number of multiples increases, the errors in model (and in turn in synthetics) will also increase. We use a narrow-offset shot record in multiple attenuation, also to avoid the wide-angle refraction of the water bottom and their multiples, as the current methodology for free-surface multiple prediction cannot properly model the refraction multiples. As marked in Figure 6a, the most difficult part of multiple attenuation is where the refraction wave just starts appearing. Figure 6b displays a shot record and the one after free-surface multiple attenuation, using a multiple prediction through inversion (MPI) method (Wang, 2004, 2007). The real shot record is generated by a point source, but Figure 6c is an equivalent line-source shot gather, after partial compensation as follows.

Before input to waveform tomography, a shot record of real seismic data needs to be partially compensated, to become a gather generated from a line source. For a two-dimensional (2-D) case, the 2-D Green's function is

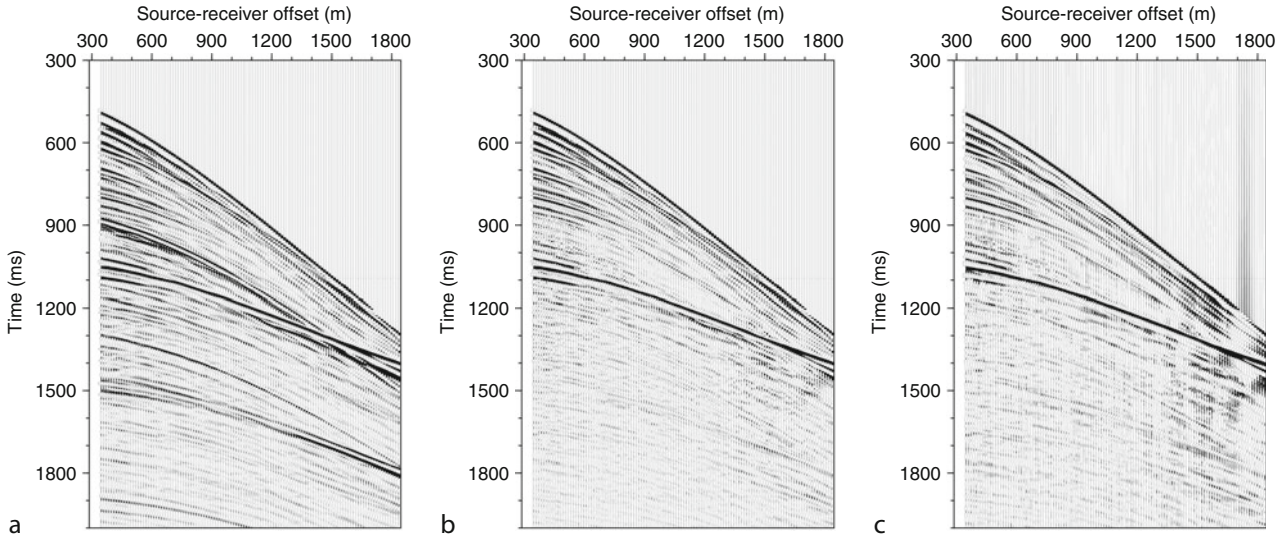
$$G_{2D}(\mathbf{r}, \mathbf{r}_s, \omega) = \left(\frac{c}{i8\pi\omega R}\right)^{1/2} \exp\left(-i\omega\frac{R}{c}\right), \tag{32}$$

where $R = |\mathbf{r} - \mathbf{r}_s|$ is the distance from the source, and c is the acoustic velocity of the medium. For a three-dimensional (3-D) case, Green's function is

$$G_{3D}(\mathbf{r}, \mathbf{r}_s, \omega) = \frac{1}{4\pi R} \exp\left(-i\omega\frac{R}{c}\right). \tag{33}$$

Comparing the 2-D and 3-D Green's functions produces a partial compensation operator as

$$W = \sqrt{\frac{2\pi R c}{i\omega}}. \tag{34}$$



Seismic, Waveform Modeling and Tomography, Figure 6 A marine seismic data example. (**a**) A sample shot record with 120 traces. (**b**) The shot record after multiple attenuation. (**c**) The same shot record after partial compensation.

In the time domain, the operator W shows the following behavior (in the far-field),

$$W(t) = D_{-1/2}(t)\sqrt{2\pi Rc}, \tag{35}$$

where $D_{-1/2}(t)$ is a half-integrator, defined as the inverse Fourier transform of $(i\omega)^{-1/2}$ (Deregowski and Brown, 1983). For a narrow-offset, reflection geometry, one can assume $2R \propto ct$ and obtain

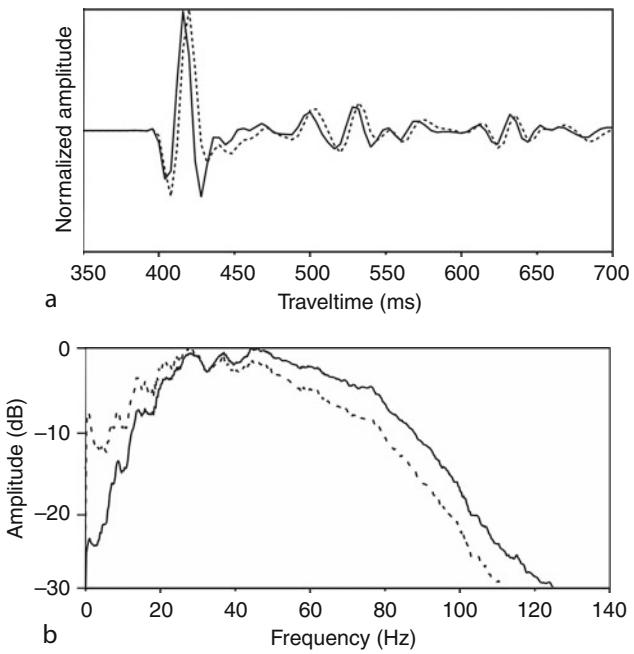
$$W(t) \propto D_{-1/2}(t)c(t)\sqrt{t}. \tag{36}$$

Therefore, this partial compensation may be implemented in two steps: scaling $c(t)\sqrt{t}$ in the time domain, and multiplication $(i\omega)^{-1/2}$ in the frequency domain.

Figure 7 closely compares a seismic trace from a point source (solid curve) and a trace after partial compensation (dash curve). With the application of the operator $(i\omega)^{-1/2}$, wavelets in the trace from an equivalent line source are broader than those actually generated from a point source.

Alternatively, one could use a 2.5-D wave modeling and inversion scheme for waveform tomography. But as it involves integration along the infinite line-source direction that is perpendicular to the source-receiver profile, a 2.5-D scheme would take a much longer running time, compared to the 2-D wave modeling and inversion (Song and Williamson, 1995).

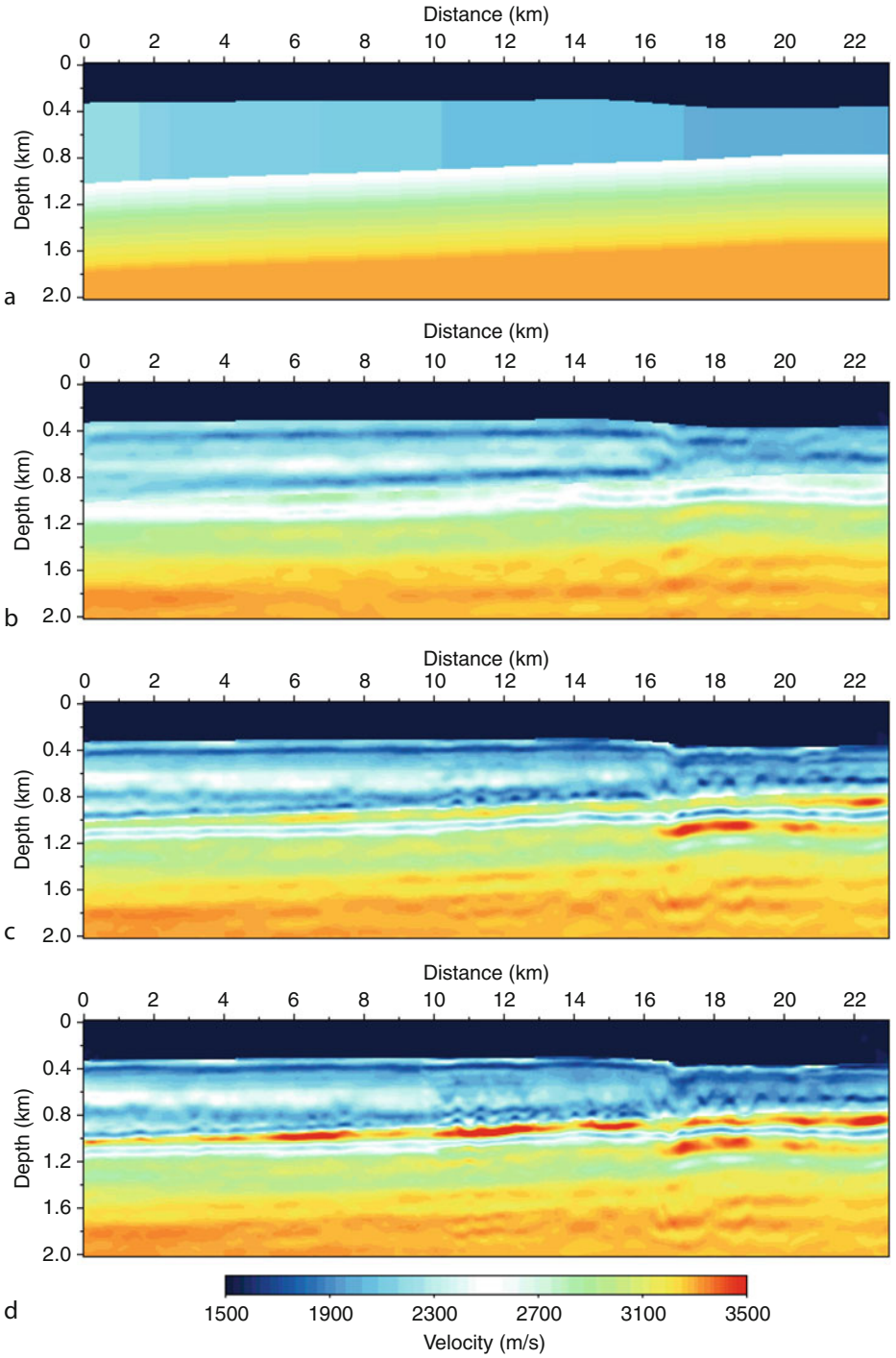
The frequency ω can also be complex-valued, to include an exponential function either for the amplitude attenuation or compensation. The attenuation used in wavefield $P(t)$ is for suppressing the wrap-around effect in the Fourier transform domain. An opposite sign with an exponential increase in $P(t)$ could also be used to boost the energy of deep reflections. In the latter case, the real data should be balanced with an automatic gain control.



Seismic, Waveform Modeling and Tomography, Figure 7 (**a**) Comparison between a seismic trace from an actual point source (*solid*) and the trace after partial compensation (*dash curve*). Wavelets in a trace from a line source (i.e., after partial compensation) are broader than those from a point source. (**b**) Comparison of amplitude spectra of seismic traces from a point source and a line source.

Reflection-seismic waveform tomography

In the reflection-seismic data, there is no significant energy recorded at low frequencies less than 6 Hz. A travel time inversion is often performed to generate the initial velocity model for the iterative waveform tomography (Pratt et al., 2002). Referring to the stack section (Figure 5), we pick the travel times of two reflections from the pre-stack seismic data, and run a travel time



Seismic, Waveform Modeling and Tomography, Figure 8 Waveform tomography. (**a**) The initial velocity model built from travel time tomography. (**b**) The velocity model of waveform tomography using frequencies in the range of 6.9–7.5 Hz. (**c**) The velocity model of waveform tomography using frequencies in the range of 6.9–13.8 Hz. (**d**) The final velocity model obtained from the waveform tomography using all frequencies in the range of 6.9–30 Hz.

inversion to generate a layered velocity model with two interfaces (Figure 8a). The first layer is a water layer with velocity 1,500 m/s. The second layer has a linear velocity of 2,200 m/s at one end and 2,000 m/s at distance 23 km.

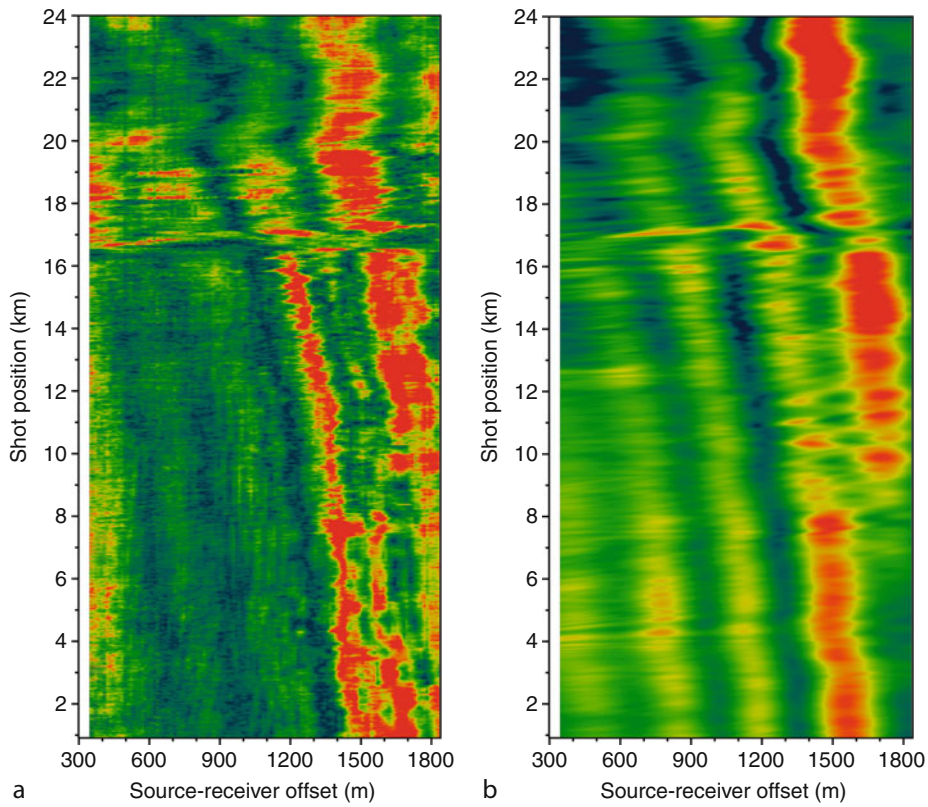
In the frequency-domain waveform tomography, we use a group of three frequencies simultaneously in each iterative inversion. Using a group of neighboring frequencies in the input can suppress the noise effect in the real data, and also more data samples used in an inversion means a much better determined inverse problem. In this example, there are 26 groups of frequencies in total in the range of 6.9–30 Hz with an interval of 0.3 Hz. The first group includes frequencies 6.9, 7.2, and 7.5 Hz, and the last group includes frequencies 29.4, 29.7, and 30 Hz. Figures 8b and 8c are the velocity models from waveform tomography using frequencies in ranges of 6.9–7.5 Hz and 6.9–13.8 Hz, respectively, where Figure 8d is the final velocity model obtained from waveform tomography using all frequencies in the range.

Although the starting model generated by travel time tomography generates a smooth boundary for the water bottom, waveform tomography produces a sharp geometry with a spatial variation close to that shown in seismic sections (Figure 5). The tomography image shows clearly a stratified structure underneath the water layer. In the second layer, at the distance between 0 and 10 km, there is a high-low-high vertical velocity variation immediately underneath the water bottom, between depth of 350 and 700 m at the left end of the profile, and a high-low-high-low vertical variation above the second interface. At the distance between 17 and 23 km, there is also a low-high-low vertical velocity variation within the second layer.

Most significantly, in the third layer, high-low-high velocity channels appear immediately underneath the second interface and cross the entire section. These thin-layers separation could not be generated by any conventional travel time tomography. Beneath that, the velocity pattern in the depth between 1.2 and 1.6 km varies laterally between 0–5, 5–17, and 17–23 km in distance.

Seismic reflection is generated due to the contrast in impedance, which is the product of density and velocity. In the inversion example here however, the density is assumed to be constant. Therefore, for a more accurate quantitative interpretation of the high velocity variation within the topmost thin layer immediately underneath the water layer, a density contrast between the water and water-bottom sediment needs be compensated.

Figure 9 compares a sample frequency slice (at 9.3 Hz) of real data input to waveform tomography and synthetic data generated from the inversion result. The vertical axis is the shot position in the surface, and the horizontal axis is the source-receiver offset. The data displayed are the



Seismic, Waveform Modeling and Tomography, Figure 9 (**a**) A sample amplitude slice of real seismic data at frequency 9.3 Hz. (**b**) The amplitude slice of synthetics at the same frequency, obtained from the waveform tomography result.

amplitudes of complex-valued samples in the frequency domain, and are normalized based on their RMS values. For the far-offset seismic traces, there are strong amplitudes for post-critical angle reflections, which play an important role in the inversion for the shallow part of model. There is also a good fit at near-offset traces, which influence the high wave number perturbations of the velocity field.

Summary

Waveform tomography uses the original waveform recordings and wave equation modeling to extract high-resolution tomographic images from seismic data. Travel time inversion may be used first to provide a reliable initial model for the waveform inversion. The latter uses a group of frequencies simultaneously in an iterative inversion, and proceeds from low to high frequencies. Even for reflection-seismic data with limited source-receiver offset, the waveform tomography also can potentially image subsurface features with detailed spatial variation at sub-wavelength scales.

Bibliography

Alford, R. M., Kelly, K. R., and Boore, D. M., 1974. Accuracy of finite-difference modelling of the acoustic wave equation. *Geophysics*, **39**, 834–842.

Boonyasiriwat, C., Valasek, P., Routh, P., Cao, W., Schuster, G. T., and Macy, B., 2009. An efficient multiscale method for time-domain waveform tomography. *Geophysics*, **74**, WCC59.

Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G., 1995. Multiscale seismic waveform inversion. *Geophysics*, **60**, 1457–1473.

Deregowski, S. M., and Brown, S. M., 1983. A theory of acoustic diffractors applied to 2-D models. *Geophysical Prospecting*, **31**, 293–333.

Kelly, K. R., Ward, R. W., Treitel, S., and Alford, R. M., 1976. Synthetic seismograms: a finite difference approach. *Geophysics*, **41**, 2–27.

Lailly, P., 1984. Migration methods: partial but efficient solutions to the seismic inverse problem. In Santosa, F., Pao, Y. H., Symes, W., and Holland, Ch. (eds.), *Inverse Problems of Acoustic and Elastic Waves*. Philadelphia, PA: SIAM.

Levander, A., 1988. Fourth-order finite-difference P-SV seismograms. *Geophysics*, **53**, 1425–1437.

Pratt, R. G., 2008. Waveform tomography – successes, cautionary tales, and future directions. In workshop 11, "Full Waveform Inversion – Current Status and Perspectives," *70th EAGE Conference & Exhibition*, Rome.

Pratt, R. G., and Shipp, R. M., 1999. Seismic waveform inversion in the frequency domain, Part 2: fault delineation in sediments using crosshole data. *Geophysics*, **64**, 902–914.

Pratt, R. G., and Worthington, M. H., 1990. Inverse theory applied to multisource crosshole tomography. Part I: acoustic wave-equation method. *Geophysical Prospecting*, **38**, 287–310.

Pratt, R. G., Shin, C., and Hicks, G. J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, **133**, 341–362.

Pratt, R. G., Gao, F., Zelt, C., and Levander, A., 2002. A comparison of ray-based and waveform tomography: implications for migration. In *64th EAGE Conference & Exhibition*, Florence.

Rao, Y., and Wang, Y., 2009. Fracture effects in seismic attenuation images reconstructed by waveform tomography. *Geophysics*, **74**, R25–R34.

Sirgue, L., and Pratt, R. G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies and waveform inversion. *Geophysics*, **69**, 231–248.

Song, Z. M., and Williamson, P. R., 1995. Frequency-domain acoustic wave modeling and inversion of crosshole data, I: 2.5-D modeling method. *Geophysics*, **60**, 784–795.

Štekl, I., and Pratt, R. G., 1998. Accurate viscoelastic modeling by frequency-domain finite differences using rotated operators. *Geophysics*, **63**, 1779–1794.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation:. *Geophysics*, **49**, 1259–1266.

Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation*. Amsterdam: Elsevier.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. *Geophysics*, **51**, 889–901.

Wang, Y., 2004. Multiple prediction through inversion: a fully data-driven concept for surface-related multiple attenuation. *Geophysics*, **69**, 547–553.

Wang, Y., 2007. Multiple prediction through inversion: theoretical advancements and real data application. *Geophysics*, **72**, V33–V39.

Wang, Y., and Rao, Y., 2006. Crosshole seismic waveform tomography, I: Strategy for real data application. *Geophysical Journal International*, **166**, 1237–1248.

Wang, Y., and Rao, Y., 2009. Reflection seismic waveform tomography. *Journal of Geophysical Research*, **114**, B03304.

Warner, M., Stekl, I., Umpleby, A., Morgan, J., Pain, C., and Wang, Y., 2008. 3D wavefield tomography: problems, opportunities and future directions. In *70th EAGE Conference & Exhibition*, Rome.

Wu, R., and Toksöz, M. N., 1987. Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, **52**, 11–25.

Cross-references

Seismic Data Acquisition and Processing
Seismic Imaging, Overview
Seismic, Ray Theory
Traveltime Tomography Using Controlled-Source Seismic Data

SEISMICITY, INTRAPLATE

Paul Bodin
Pacific Northwest Seismic Network (PNSN), University of Washington, Seattle, WA, USA

Definition

Intraplate Seismicity refers to earthquakes that take place on faults other than those that form the boundaries between Earth's tectonic plates.

Historical context

From the inception of plate tectonics as a simple set of unifying concepts, Earth scientists recognized that the notion of completely rigid plates should not be taken to be rigorous or complete. Deformations of Earth's crust and upper mantle, and the earthquakes that generally accompany them, must take place off of recognized plate interfaces. This fact was evident theoretically because in order for

plate tectonics to operate on timescales greater than one Wilson cycle, new plate boundaries must form within plates, so deformation must take place off of existing plate boundaries. In addition, it was unassailable observationally because seismicity off of recognized plate boundaries clearly *does* take place, albeit at a relatively trivial pace compared to Earth's total earthquake production.

Two generations later, despite the fact that they obviously exist, intraplate earthquakes remain enigmatic and challenging; fitting subjects for special study. The dominant questions about these earthquakes not only reveals our incomplete understanding of earthquake physics in general, but also highlights a number of issues specific to active tectonics and seismic hazard in intraplate settings.

Definitions and classification

We may define interplate earthquakes as being those that rupture parts of faults that connect directly mechanically to the globe-girdling system of recognized boundaries between tectonic plates. Intraplate earthquakes break faults that do not so connect. The difference is one of ultimate slip. Because interplate faults form a network within which new plates may be created and old plates consumed, slip on them is unbounded. On the other hand, the total slip on intraplate faults must terminate at the ends of the fault. So increasing slip comes at the cost of increasing strain in the surrounding rocks which, itself being presumably limited, caps how much any given fault can slip. Therefore, intraplate faults may be expected to slip generally more slowly, and less far, than interplate faults, and this possibly to have some consequence on the characteristics of earthquakes on the two types of faults.

It is customary for Earth scientists to distinguish further between intraplate earthquakes that break faults within *mobile zones* and those that take place within *stable zones*. Mobile zones are regions that, although not directly on recognized plate boundary fault, have been experiencing regional deformation distributed over a broad area near a plate boundary. An example is the active faulting across China, Tibet, and Mongolia, the deformations of which may be clearly related to three-dimensional complexities in the plate boundary framework. Stable zones are not only sufficiently remote from plate boundaries to be insensitive to their geometric complexities, but also are stable in the sense that significant tectonic events have not affected them during the Cenozoic Era ($\sim$66 million years). They are generally expected to be areas of cold, thick continental lithosphere. Earthquakes in stable continental regions, often designated SCR earthquakes, are particularly enigmatic.

Reasons to distinguish intraplate earthquakes

One reason to study intraplate earthquakes (and particularly SCR earthquakes) as a class involves a practical consequence: they strike the most unwary victims. Because of the long recurrence intervals between intraplate earthquakes, the threatened populace is generally unprepared and the structures they build are particularly vulnerable. It has recently been noted that this argument is becoming less restrictive to intraplate earthquakes as population growth within plate boundary regions has taken place so quickly of late that even these regions are becoming less prepared as a whole (Bilham, 2009). Nevertheless, the argument remains applicable in plate interiors where the building stock may be not only not up to modern seismic codes, but is generally older.

In part, SCR earthquakes remain especially enigmatic not merely because they are rare, but also because their recurrence may be so long that (1) the sizes and sometimes locations of previous earthquakes are not well known, even for those that have occurred in historical times, and (2) it is usually not possible to ascertain what their recurrence is with certainty. This leads to great uncertainty about the seismic hazards in intraplate regions. Is damaging earthquake shaking possible anywhere (albeit with small probability within any 50-year time period)? Or are even SCR earthquakes restricted to occur in places with discernible (and hence predictable) characteristics? This is tantamount to a question of whether the entirety of Earth's crust at critical stress levels such that only small fluctuations or differences in fault strength control the occurrence of earthquakes.

With such long recurrence intervals, seismological observations are sparse and we rely on the paleoseismic evidence contained within the geological record to try to understand them. It is clear that plate tectonics is accurate to first order and that intraplate deformation takes place very slowly, which makes them a challenge for even modern high-precision geodetic techniques. It seems likely that progress in understanding intraplate earthquakes will remain slow for the time being.

Another reason they are of special interest is that they may reveal clues about elusive earthquake physics. While Earth scientists believe that nearly all tectonic earthquakes are caused by sudden slip on a fault driven by stresses in the surrounding rock (the elastic rebound theory), the specific mechanisms that localize rupture initiation and growth of a rupture remain obscure. Systematic differences between fault, stress, and geological characteristics of intraplate faults discussed below may be exploited to test hypotheses of the important controls of earthquake fault rupture.

Do intraplate and interplate earthquakes scale differently?

Because intraplate earthquakes, and particularly those in stable zones, have longer recurrence intervals, it is often hypothesized that their host faults will have longer time to "heal" between earthquakes (e.g., Scholz, 2002). Fault healing here refers to a set of hypothesized fault processes such as the formation of minerals that bind the sides of the

fault together making them stronger over time. Thus, intraplate earthquakes may be expected to break under higher loads – at higher stress levels – than their cousins that break more active plate boundary faults. The anticipated result is that they "scale" differently: that is, big earthquakes differ from small earthquakes in systematically different ways in interplate as opposed to intraplate regions. Specifically, intraplate earthquake ruptures will slip more in an earthquake per unit of fault area, relieve more stress (have a larger "stress drop") and produce more seismic energy for their size, than the earthquakes we are more familiar with at plate boundaries. Thus seismologists expect, and arguably observe, that intraplate earthquakes tend to have higher stress drops than interplate earthquakes. Seismological evidence suggests that as a class their stress drops may be a factor of 2–5 higher than plate boundary earthquakes (Kanamori and Allen, 1986). But because observations are so few, much uncertainty remains about this issue.

Source scaling studies are further hindered by systematic differences between seismic wave propagation in intraplate regions and that in more mobile or active zones. Because rock surrounding plate boundaries is often less intact and hotter than in plate interiors, seismic wave energy is more strongly absorbed, or attenuated, than in the cold hard hearts of plates. So, per unit of energy released at their source, the effects of intraplate earthquakes are more widely distributed – they have a longer reach.

What drives intraplate earthquakes?

Probably the most critical enigma of intraplate seismicity concerns the source of driving stresses. It is easy to invoke "left over" or residual stresses originating from geometric irregularities or variations in frictional properties on the master plate boundary faults to explain the origin of mobile zone intraplate earthquakes. Remote from plate boundaries, occasionally sources of localized stresses may be recognized. One clear example is from intraplate regions experiencing uplift from glacial rebound, as ice sheets retreat with global warming. This mechanism well explains the seismicity pattern of Sweden, for example (e.g., Mörner, 2009). Swarms of seismicity often accompany geothermal activity, even at sites in the remote interior of tectonic plates (e.g., Ibs-von Seht et al., 2008). Where such sources of localized stress are not evident, however, it is more problematic to assign the provenance of seismic stress. Modeling studies suggest that stresses within plates may be best explained as a combination of resolved stresses originating at plate boundaries (although they may be thousands of kilometers distant) plate bending due to a plate's travel over a non-spherical Earth, or forces arising from convective patterns in Earth's mantle.

Many scientists presume that intraplate earthquakes take place at sites where crustal rocks have been significantly weakened compared to those in surrounding rocks. In this model, seismicity should be associated with reactivated fault zones. Big intraplate earthquakes are thought to be restricted to areas that have previously been faulted, say within failed continental rift systems (e.g., Johnson, 1996).

One intriguing idea is that the strength of faults could be reduced by high pore-fluid pressure. Fluids (which could include water or even carbon dioxide or hydrocarbons) that occupied the internal portion of a fault could become pressurized and reduce the frictional resistance of the fault to slipping. That would imply that seismically active faults in intraplate regions are those that contain significant quantities of fluids at high confining stress. Sufficient stress to drive earthquakes could indeed be present everywhere within plate interiors, and the conditions to release the stress controlled by something external.

Examples

The moment magnitude (Mw) = 6.8 "Nisqually" earthquake of January 23, 2001 in Washington state, USA, provides an example of an intraplate earthquakes breaking a fault within the crust of the subducting ocean plate (Ichinose et al., 2006). The earthquake resulted from slip on a normal fault driven by bending stresses as the downgoing slab flexes in its descent. On a larger scale, the Cascadia subduction zone provides an interesting example of the interplay between inter- and intraplate earthquakes in that while the master plate boundary megathrust fault is broken about every 500 years with large subduction thrust events (e.g., Satake and Atwater, 2007), there appears to be no current seismicity on the plate boundary fault. Rather all earthquakes either historical or instrumentally recorded are either intra-slab events or crustal earthquakes within the overriding North America plate. These events are interplate by our definition. However, they are clearly associated with a nearby subduction interface and are driven by residual stresses that the plate boundary is unable to completely relieve.

The May 13, 2008 Mw = 7.9 Wenchuan earthquake in the Szechuan Province of China is an earthquake in the tectonically mobile zone of Asia, and is part of the escape of material from Asia and the uplifting of the Himalaya due to the collision of the Indian subcontinent with Asia. About 300 km of a fault on the boundary between the Longmenshan orogenic belt and the Szechuan basin broke in the earthquake (Burchfiel et al., 2008). With a high overall stress drop of 17 MPa and localized patches of up to 53 MPa (Zhang et al., 2009), the Wenchuan earthquake had large slip and disastrous effects on a populace that was largely unaware of the potential for, and unprepared for, this long-recurrence-interval (2,000–10,000 years) intraplate earthquake.

The January 26, 2001 Mw = 7.6 Bhuj earthquake took place ~400 km? from the rather diffuse onshore portion of the left-lateral strike-slip Makran boundary between India and the Arabian and Eurasian plates, but within the

western part of the peninsular Indian shield. It broke the deeper part (~5 to ~40 km) of a previously unknown reverse fault. It was located within ~100 km of a similar-sized (though shallower) earthquake in 1835 that generated a surface rupture – the Allah Bundh in the Great Rann of Kachchh. The stress drop was a high, but not extreme ~16–20 MPa, but the earthquake was very powerful, with the source lasting only about 7 s. Widespread liquefaction and secondary effects were observed, but the large death toll (~30,000) was apparently not the result of extreme ground motions, which could have been engineered for, but rather of poor construction and preparation.

Controversy

One particularly illustrative and significant example of SCR interplate seismicity concerns the New Madrid Seismic Zone that threatens a 7-state region surrounding the lower Mississippi River valley in the south central USA (e.g., Johnson and Schweig, 1996). The region was the site of at least three major earthquakes during the winter of 1811–1812. While at the time thinly populated and rural, a return of these earthquakes today would affect ten million people and put at risk numerous cities, commerce, and transportation and communication infrastructure. The 1811–1812 earthquakes were accompanied by a host of surface effects (liquefaction, landslides, surface deformation) that remain visible in the landscape today. Ground motions from the earthquakes were also reported to be felt as far away as Charleston, South Carolina (~950 km) and even Boston, Massachusetts (~1,600 km). There is a complex zone of active microearthquakes producing several $M < 3$ earthquakes each week. Early research based on ground-shaking intensities derived from historical reports proposed that the three largest earthquakes in the series exceeded M8. The ~100-km long zone of active microseismicity, which many presume to represent the active faults, seems much too small to host even a single M8 earthquake, let alone three, unless stress drops, and therefore slip per unit length, were extremely large. The earthquakes occupy a shallow basin formed atop a failed continental rift, formed in latest Precambrian time and briefly reactivated (but once again failed) in the Cretaceous. The faults are oriented such that they are nearly optimally oriented to be reactivated as thrust and strike-slip fault in the current compressional mid-plate stress regime.

Subsequent reexamination of the historic intensity data, and taking into account the feeble attenuation of Seismic waves within rocks in the plate interior, have yielded generally smaller estimates of magnitudes, as low as M7 for the three largest 1811–1812 earthquakes. In the absence of new information, however, it might be best to admit that while these earthquakes were undoubtedly large, their magnitudes will remain forever moot. Paleoseismic studies of earthquake-induced liquefaction features in surface deposits reveal at least four repeated episodes of strong shaking (as strong as what took place in the early nineteenth century), often in clusters like the 1811–1812 events, with an average recurrence interval of ~500 years. The geological evidence for recurrent large earthquakes is balanced by the lack of primary surface faulting, or even the prominent topographic relief one would expect if much slip had taken place at the rates paleoseismic evidence infers for the past several thousand years. So either the big events are really a good deal smaller than the largest estimates of their size or seismic activity in the zone was initiated only a few thousand years ago.

Geodetic evidence, from repeated campaign surveys and continuous GPS studies show that deformation surrounding the active seismic zone (Smalley et al., 2005) is very slow, with some (e.g., Newman et al., 1999) arguing that it is not different from zero. However, given the lack of a mid-plate fault model it is not clear what deformation pattern one should expect. So there is no general agreement about what the lack of observed active deformation means, and the interpretations remain contentious. Some suggest that either the largest earthquakes to be expected from the seismic zone are not so terrifyingly large as previously thought, or perhaps that, however large they were, they are not being recharged with stress and are not to be feared. Others hold that the paleoseismic, historical, and microseismic evidence point to the likelihood of future damaging earthquakes just like the historical ones, albeit driven by processes that are still beyond our ken.

Summary or conclusions

Intraplate seismicity is a catchall phrase that encompasses a wide variety of seismogenic phenomena; essentially all earthquakes that do not occur on plate bounding faults. Locales hosting intraplate earthquakes range from broad deformation zones that surround some plate boundaries to areas that lie within the cores of continents. Larger earthquakes – especially those in what are otherwise tectonically stable cold continental crust – appear to be associated with faults that formed in an earlier tectonic setting but are reactivated by the stress field applied currently. Not all faults are so reactivated, so it is likely that some as-yet-unclear weakening mechanism causes only certain faults to fail. Localized deformation, say glacial rebound or geothermal stresses can be responsible for interplate earthquakes, but does not explain all of them. As a class of faults, intraplate faults usually slip more slowly and less far than interplate faults, and this leads to longer intraplate earthquake recurrence intervals. Also, and probably related to this, intraplate earthquakes may generally be expected to have larger stress drops. Although less frequent than plate boundary earthquakes, intaplate seismicity is very dangerous because the population and infrastructure at risk from them is less prepared than in more seismically active areas. The study of intraplate seismicity is an endeavor that, because of its low rate is best done by integrating across different source regions and combining different Earth science disciplines.

Bibliography

Bilham, R., 2009. The seismic future of cities. *Bulletin of Earthquake Engineering*, doi:10.1007/s10518-009-9147.

Burchfiel, B. C., Royden, L. H., van der Hilst, R. D., Hager, B. H., King, R. W., Li, C., Lü, J., Yao, H., and Kirby, E., 2008. A geological and geophysical context for the Wenchuan earthquake of 12 May 2008, Sichuan, People's Republic of China. *Geological Society of America Today*, **18**(7), doi:10.1130/GSATG18A.1.

Ibs-von Seht, M., Plenefisch, T., and Klinge, K., 2008. Earthquake swarms in continental rifts – a comparison of selected cases in America, Africa and Europe. *Tectonophysics*, **452**(1–4), 66–77, doi:10.1016/j.tecto.2008.02.008. ISSN 0040-1951.

Ichinose, G. A., Thio, H. K., and Somerville, P. G., 2006. Moment tensor and rupture model for the 1949 Olympia, Washington, earthquake and scaling relations for Cascadia and Global Intraslab Earthquakes. *Bulletin. Seismological Society of America*, **96**(3), 1029–1037, doi:10.1785/0120050132.

Johnson, A. C., 1996. Seismic moment assessment of earthquakes in stable continental regions II: historical seismicity. *Geophysical Journal International*, **125**(3), 639–678.

Johnson, A. C., and Schweig, E. S., 1996. The enigma of the New Madrid Earthquakes of 1811–1812. *Annual Review of Earth and Planetary Sciences*, **24**, 339–384.

Kanamori, H., and Allen, C. R., 1986. Earthquake repeat time and average stress drop. In Das, S., et al. (eds.), *Earthquake Source Mechanics*. Washington: American Geophysical Union. AGU Geophysics Monograph, Vol. 37, pp. 227–236.

Mörner, N-A., 2009. Late Holocene earthquake geology in Sweden, *Geological Society, London*, Special publications, 316, pp. 179–188, doi:10.1144/SP316.11

Newman, A., Stein, S., Weber, J., Engeln, J., Mao, A., and Dixon, T., 1999. Slow deformation and lower seismic hazard at the New Madrid Seismic zone. *Science*, **284**, 619–621, doi:10.1126/science.284.5414.619.

Satake, K., and Atwater, B. F., 2007. Long-term perspectives on giant earthquakes and tsunamis at subduction zones. *Annual Review of Earth and Planetary Sciences*, **35**, 349–374, doi:10.1146/annurev.earth.35.031306.140302.

Scholz, C. H., 2002. *The Mechanics of Earthquakes and Faulting*. Cambridge, UK: Cambridge University Press.

Smalley, R., Ellis, M. A., Paul, J., and Vanarsdale, R., 2005. Space geodetic evidence for rapid strain rates in the New Madrid seismic zone of central USA. *Nature*, **435**, 1088–1090, doi:10.1038/nature03642.

Zhang, Y., Feng, W., Xu, L., Zhou, C., and Chen, Y., 2009. Spatio-temporal rupture process of the 2008 great Wenchuan earthquake. *Science in China. Series D: Earth Sciences*, **52**(2), 145–154.

Cross-references

Earthquake, Focal Mechanism
Earthquake, Magnitude
Earthquakes and Crustal Deformation
Earthquakes, Strong-Ground Motion
Geodesy, Figure of the Earth
Geodesy, Ground Positioning and Leveling
Geodynamics
Mantle Viscosity
Paleoseismology
Plate Driving Forces
Plate Motions in Time: Inferences on Driving and Resisting Forces
Plates and Paleoreconstructions
Plate Tectonics, Precambrian

SEISMICITY, SUBDUCTION ZONE

Akira Hasegawa
RCPEV, Graduate School of Science, Tohoku University, Sendai, Miyagi, Japan

Definition

Subduction zone. Consuming plate boundary where the oceanic plate subducts into the mantle beneath the continental plate.

Interplate earthquakes. Earthquakes that occur along the plate boundary. In case of the subduction zone, they occur at the shallow portion of the boundary between the subducting plate and the overriding plate.

Intraslab earthquakes. Earthquakes that occur within the subducted plate, that is, within the slab.

Shallow earthquakes. Earthquakes that occur at depths shallower than 60 km. Focal depths of interplate earthquakes in subduction zones are usually shallower than 60 km, and so they are shallow earthquakes.

Intermediate-depth earthquakes. Earthquakes that occur at depths of 60–300 km.

Deep earthquakes. Earthquakes that occur at depths deeper than 300 km. Both intermediate-depth and deep earthquakes occur only within the subducted plate, and so they are intraslab earthquakes.

Introduction

There are seismically active and inactive areas on the Earth. Most of Earth's seismic activity is concentrated in narrow continuous belts that wrap around it. Plate tectonics proposes that the Earth's entire surface is comprised of somewhat more than ten large, rigid, mosaic-like plates, and that the relative movement of these plates causes tectonic activity at their boundaries. Consequently, most earthquakes are concentrated at plate boundaries, forming narrow, belt-shaped seismic zones.

Oceanic plates are generated at mid-ocean ridges by the upwelling of mantle material. Since the Earth's surface area is constant, the surface area does not expand with the generation of the oceanic plate. Instead, the same amount of material moves back down into the mantle. The generation of plates is balanced by the consumption of other plates elsewhere at consuming plate boundaries. *Subduction zones* are where oceanic plates plunge downward into the mantle. At subduction zones, two plates collide with each other; the heavier plate is overridden by the lighter one and sinks down into the mantle. Thus, the heavier oceanic plate subducts beneath the lighter continental plate. The subduction of the oceanic plate causes high seismic activity. In fact, it is in these subduction zones that most of the world's large earthquakes occur.

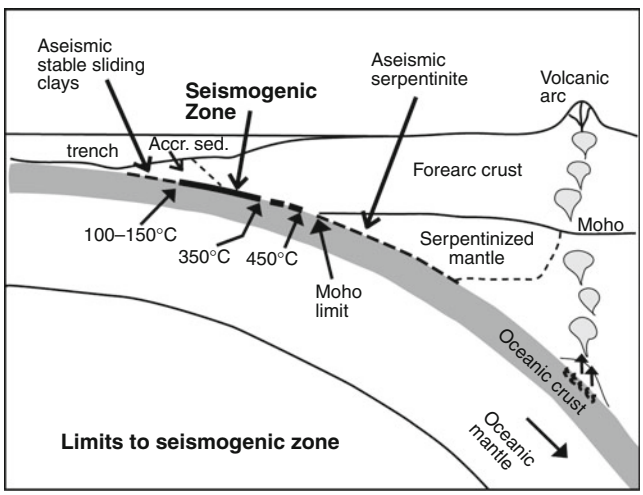
Interplate earthquakes

In subduction zones, the denser oceanic plate sinks down into the mantle underneath the lighter continental plate.

The shallow portion of the plate boundary is locked by friction, and in order to release accumulated stress, occasional slips at the plate boundary occur and cause earthquakes. Most of the world's large, shallow earthquakes are interplate earthquakes caused in this way. Interplate coupling occurs only within a specific range of depths, which seems to be primarily determined by the temperature of the plate interface.

Oleskevich et al. (1999) estimated the thermal structures of plate interfaces in several subduction zones, and compared them with the updip and downdip depth limits of source areas of large interplate earthquakes. The results show: (1) the updip limit of the depth of interplate earthquakes corresponds with the depth at which the temperature of the plate boundary reaches 100–150°C and (2) the downdip limit approximately matches the depth at which the temperature reaches 350°C (Figure 1). In the shallowest part of plate interfaces, smectite clays, which are clay minerals distributed in the sediments there, prevent interplate coupling by producing stable sliding. The temperature rises with increasing depth and when it reaches 100–150°C, dehydration decomposition occurs and the smectite clay turns into illite and chlorite. This transformation is considered the onset of interplate coupling.

Then, deeper, where the temperature exceeds 350°C, interplate coupling ceases and stable sliding takes place again. Figure 1 shows that the transition zone from unstable sliding to stable sliding occurs at depths where the temperature ranges from 350°C to 450°C. In the case where the depth at which the temperature of the plate interface reaches 350°C is deeper than the depth at which it contacts the Moho of the upper plate, the contact zone with the Moho is considered to be the downdip limit. The reason the downdip is limited to the Moho contact zone is that stable sliding prevails where the mantle of the upper plate contacts the oceanic plate because the mantle may be serpentinized by water supplied by the dehydration of the subducting oceanic plate. It is considered that interplate earthquakes do not occur any more under those conditions (Hyndman et al., 1995). In some subduction zones, such as in northeast Japan, the upper plate's mantle is not serpentinized, since the seismic wave velocities there show normal values of non-serpentinized mantle materials. In such cases, the downdip limit is once again determined by temperature, although the specific temperature at which the transition occurs is not yet known.

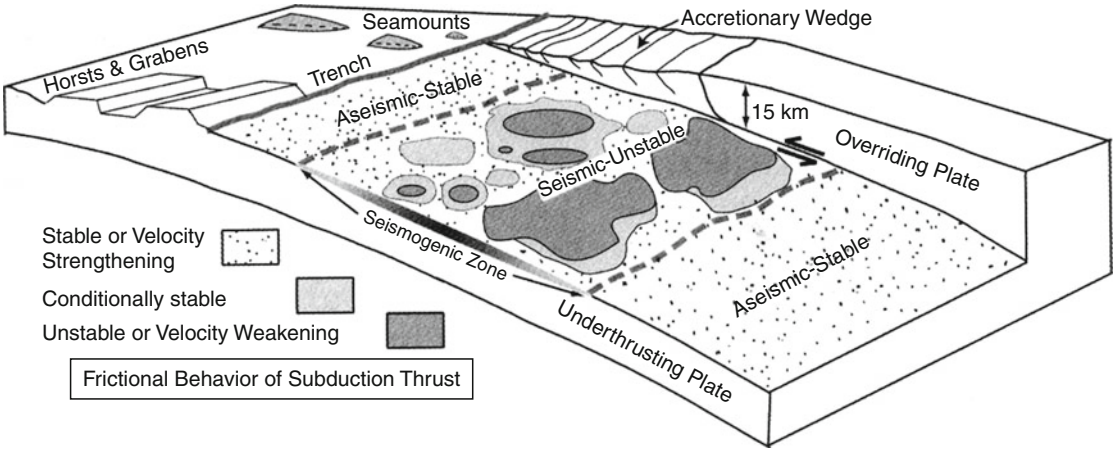


Seismicity, Subduction Zone, Figure 1 Sketch of the subduction plate boundary seismogenic zone (Oleskevich et al., 1999, copyright by the American Geophysical Union).

The degree of interplate coupling can be indicated by how much interplate slip is accounted for by earthquakes. This index is called "seismic coupling" and denotes the percentage of the seismic slip among all slip amounts occurring at the "seismogenic zone" in Figure 1. If seismic coupling is 1, all slip occurring in that area is caused by earthquakes. In contrast, if it is 0, it means that all slip occurring in that area is aseismic stable sliding. Seismic coupling differs from one subduction zone to another, and the larger the maximum magnitude of interplate earthquakes is in a subduction zone, the larger is its seismic coupling (Lay and Kanamori, 1981).

The younger the age of the subducting plate and the faster the plates are converging, the larger the maximum earthquake magnitude (Ruff and Kanamori, 1980) and so the stronger is the interplate coupling. A faster convergence velocity results in a higher rate of stress accumulation, which would accompany a higher rate of earthquake activity. Young plates are generally less dense than older plates, and being relatively light, the gravitational force pulling them downward is weaker and their subduction angle is lower than with older, denser plates. When the subduction angle is low, the contact area between the subducting plate and the upper plate is large. The large contact area and relatively weak downward force produce strong interplate coupling between the plates. Moreover, younger plates have relatively smoother plate surfaces because they do not have long alteration histories after their generation at the midoceanic ridge, which probably allow the formation of larger focal areas of large maximum earthquakes. For these reasons, if two plates have the same subduction velocity, the younger plate will have stronger interplate coupling and larger maximum earthquake magnitude (Kanamori, 1977).

Interplate coupling along the seismogenic zone is heterogeneous both in space and time; the seismogenic zone is not fully locked, as mentioned above, and some portions of it will slide stably during interseismic periods. Seismic waveform inversions of large interplate earthquakes have revealed spatially heterogeneous seismic slip along the seismogenic zone, and backslip inversions of geodetic data, which estimate spatial distribution of interplate coupling, have confirmed spatially heterogeneous strain accumulation along the zone during the interseismic period. A conceptual model of frictional properties on the fault surface based on laboratory-derived fault constitutive laws



Seismicity, Subduction Zone, Figure 2 A conceptual model of frictional properties on the subduction plate boundary (originally from Bilek et al. (2002) and modified by Dixon and Moore (2007), copyright by the Columbia University Press).

(Figure 2) explains the observed heterogeneous interplate coupling along the seismogenic zone. It involves patches of frictionally unstable fault surface surrounded by stable and conditionally stable fault surfaces. Frictionally stable portions of the fault are subject to steady aseismic sliding, whereas unstable patches (asperities) remain locked during the interseismic period. Then they finally fail, as evidenced by the occurrence of earthquakes. Conditionally stable portions typically slide stably, but may fail and slide dynamically if loaded at high strain rates by the seismic slip of a neighboring asperity. Earthquakes nucleate only in the unstable patches, and once an earthquake occurs in an unstable patch its seismic slip propagates throughout the patch, with some penetration into the neighboring conditionally stable areas. This conceptual model indicates that the larger the area occupied by the frictionally unstable patches in the seismogenic zone, the larger the seismic coupling of that subduction zone is.

Shallow inland intraplate earthquakes

Convergence of two plates at subduction zones causes stress accumulation also in the overriding continental plate, producing deformation in it well beyond the trench axis. Usually, this deformation is compressional crustal shortening, which partially accommodates the plate convergence. Shallow earthquakes occur in the crust of the overriding plate, reflecting this ongoing compression. Focal mechanisms of these earthquakes are thrust fault or strike-slip fault with compressional axis oriented nearly in the plate convergence direction. Recurrence intervals of these types of earthquakes are much longer than those of interplate thrust earthquakes. However, once they occur, it often causes serious damages in lives and societies due to the shallow hypocenter and proximity to inhabited areas.

In most cases, these shallow inland earthquakes occur in the depth range between 1–3 and 10–20 km, forming the seismogenic zone mainly in the upper crust of the overriding plate. Deformation is made by aseismic ductile flow at shallow depths above the upper bound because of the presence of unconsolidated granular material, and at large depths below the lower bound because of the onset of plasticity at a critical temperature (Scholz, 1998). This critical temperature is considered to be about 300–400°C, and so the lower bound of the seismogenic zone varies depending on the local thermal gradient.

Active faults, which are surface traces of repeated fault movements by earthquakes, show locations where in the seismogenic zone of the overriding plate large shallow earthquakes have occurred in the recent past, the Quaternary Period. Surveys of active faults further provide information on history of earthquakes, including their magnitudes and the time of the last event, which is basic and important for long-term earthquake forecast of shallow inland intraplate earthquakes.

Although we can learn from active faults where large earthquakes have occurred, it is difficult to understand why earthquakes occur in those locations. Recent investigations based on dense GPS and seismic observation network data in Japan have provided some information on this problem. Analyses of GPS data showed the existence of belt-like zones with concentrated crustal shortening in the overriding plate, along which many large shallow earthquakes have occurred or active faults are distributed (Sagiya et al., 2000; Miura et al., 2004). Seismic tomography studies revealed prominent low seismic velocity zones in the lower crust to the uppermost mantle right beneath these concentrated deformation zones (Nakajima et al., 2001; Nakajima and Hasegawa, 2007). The low seismic velocities are inferred to reflect the existence of aqueous fluids, which are perhaps supplied from the subducted slabs. These observations suggest that the fluids weaken the surrounding crustal materials and cause the concentrated deformation there, leading to large shallow inland earthquakes (Iio et al., 2004; Hasegawa et al., 2005). If this is the case, places to which aqueous fluids

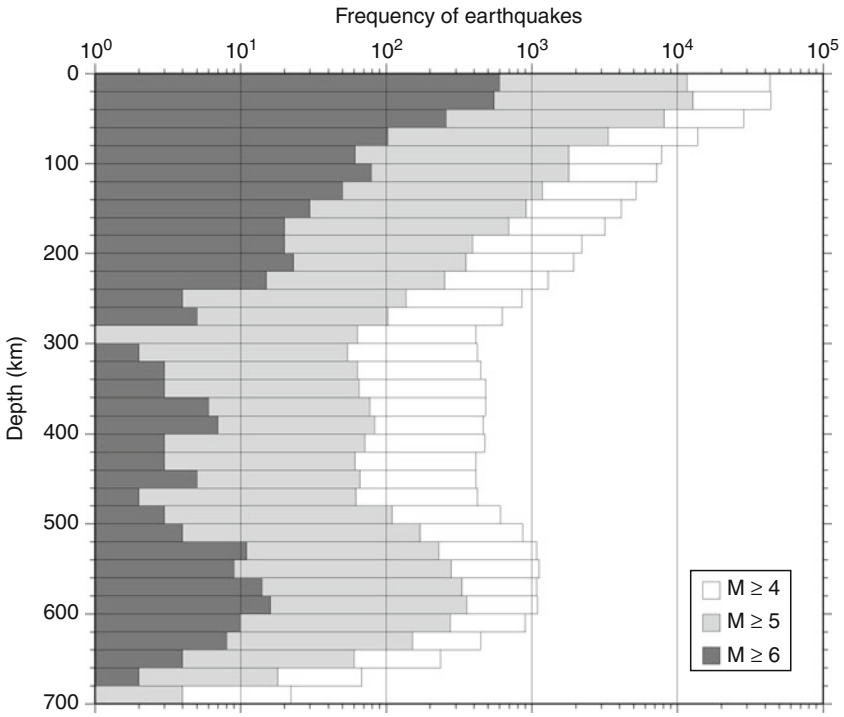
are supplied intensively from the mantle wedge below are considered to be locations where large shallow earthquakes occur.

Intraslab earthquakes

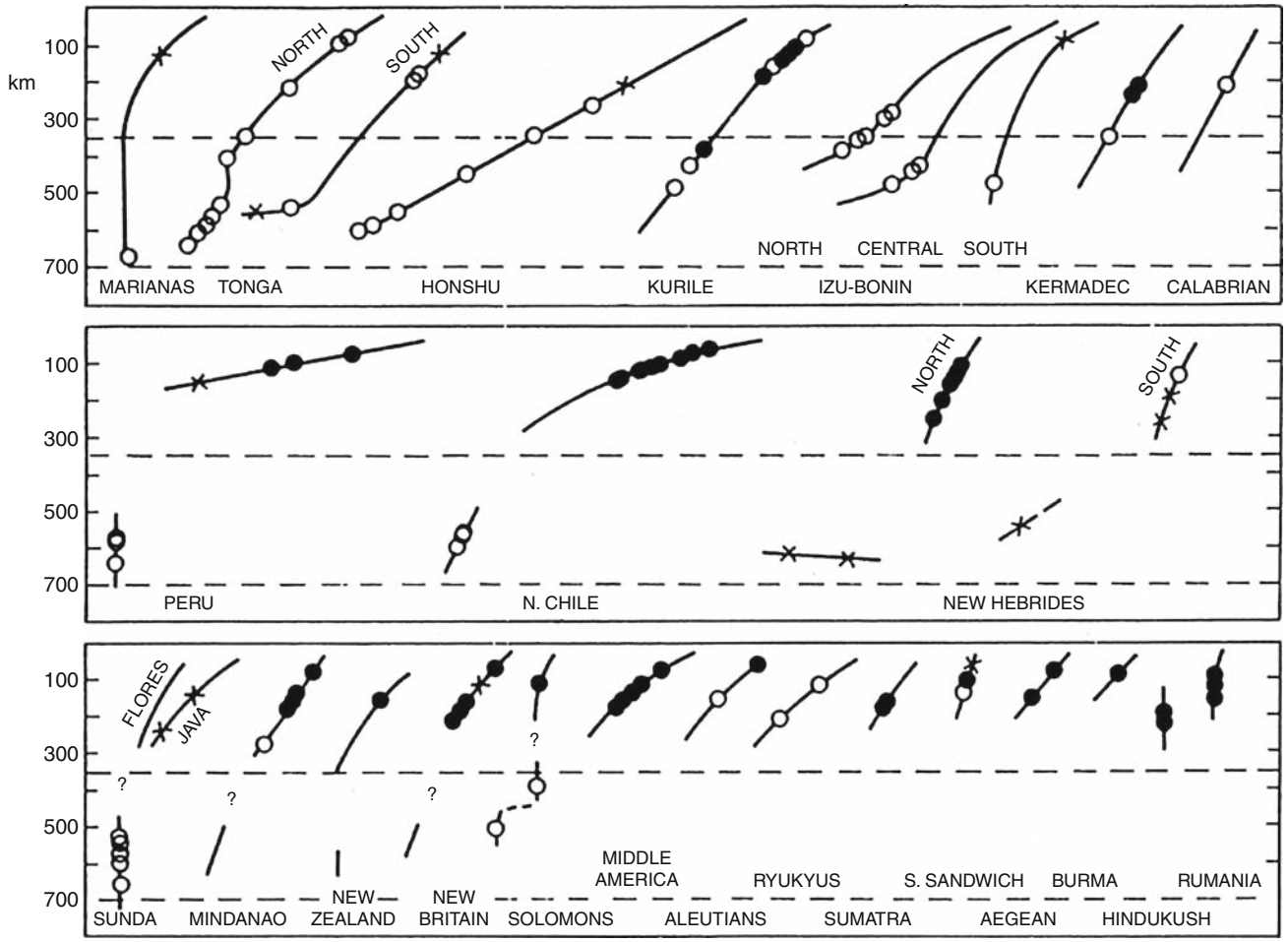
Earthquakes that occur at depths of more than about 60 km are found only in subduction zones. That is because these earthquakes occur in subducting oceanic plates or slabs. They are distributed in an inclined plane in the mantle, which is called "deep seismic zone" or "Wadati–Benioff zone," after its discoverers. The frequency distribution of earthquakes for each depth range throughout the world (Figure 3) shows that depths from 0 to 50–60 km have the greatest occurrence frequency. Proceeding deeper, the frequency decreases nearly monotonically until a depth of about 300 km, then starts to increase with depth from that point, reaching a local maximum at 500–600 km deep. The earthquake depth range can be divided into three zones: shallow (0–60 km), intermediate depth (60–300 km), and deep (deeper than 300 km). Of these, intermediate-depth earthquakes and deep earthquakes occur within the slab so they are called intraslab earthquakes (or slab earthquakes).

In many cases, the focal mechanisms of intraslab earthquakes are classified as either downdip compression (DC) earthquakes, in which the compressional axis is oriented in the dip direction of the slab, or downdip extension (DE) earthquakes, in which, on the contrary, the tensional axis is oriented in the dip direction of the slab. There is a systematic tendency in the depth distribution of focal mechanisms that depends on the age of the subducting plate. That is, as indicated in Figure 4: (1) DE-type earthquakes dominate in young plate subduction zones such as in Central America, in which intraslab earthquakes occur only to a depth of about 250 km. (2) Intraslab earthquakes occur to a depth of about 650 km in old plate subduction zones such as Tonga and the Izu-Bonin arc, and the DC-type focal mechanism dominates at all depths in those areas. (3) Subduction zones such as the Kermadec and Kuriles areas have characteristics intermediate between (1) and (2); DE-type earthquakes dominate in shallow reaches while DC-type earthquakes dominate deeper. (4) In subduction zones like Chile and the New Hebrides areas, in which intraslab earthquakes do not occur at depths between 300 and 500 km, the DE-type dominates at shallow depths while the DC-type dominates deeper.

Isacks and Molnar (1971) explain the formation of these earthquake-generating stress fields in the slab as follows. The depths to which the downdip ends of slabs reach constitute four different conditions (Figure 5). In one configuration, the end of the slab reaches the top of the high strength lower mantle (Figure 5c), which corresponds to the conditions described in (2) above. In two other configurations, it ends within the upper mantle (Figures 5a and b) corresponding to (1) and (3) above. In the fourth case, the slab is not continuous, but is divided into upper and lower segments (Figure 5d) corresponding to (4) above. When slabs subduct into the mantle, the resistive force becomes



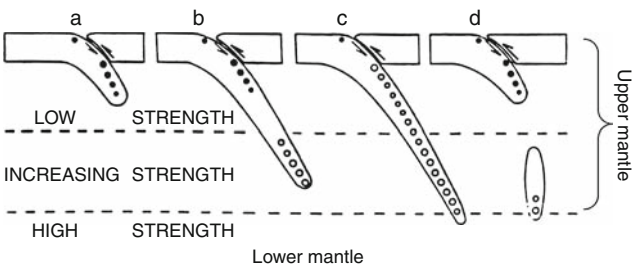
Seismicity, Subduction Zone, Figure 3 Focal depth distribution of global seismicity. Earthquakes with magnitudes greater than 4, 5, and 6 in the EHB catalog (Engdahl et al., 1998) for the period from 1964 to 2004 are plotted.



Seismicity, Subduction Zone, Figure 4 Global summary of stress states within subducting slabs (Isacks and Molnar, 1971, copyright by the American Geophysical Union). Open and solid circles show downdip compression and downdip extension, respectively. Crosses indicate that neither the compression nor tension axis is approximately parallel to the slab dip.

stronger with increasing depth, and the earthquake-generating stress fields discussed above are formed. Although the above explanation by Isacks and Molnar seems plausible at first glance, recent seismic tomographic studies are now revealing that the end of the slabs reach the top of the lower mantle in most subduction zones. Therefore, the interpretation provided in Figure 5 is not always correct.

Slab mineralogy studies have shown that depth perturbations of phase transformation boundaries resulting from thermal anomalies within the slabs also generate positive and negative buoyancy forces depending on depth, and so are major contributors to stress fields in the slabs (e.g., Bina et al, 2001). Figure 6 shows a diagram of the mineralogy in a subducting slab. Since the slab is colder than the surrounding mantle, the α-olivine to β-spinel phase transformation at 410 km depth and the β-spinel to γ-spinel phase transformation at 550 km depth become shallower, whereas decomposition of γ-spinel to perovskite + magnesiowustite at 660 km depth becomes deeper

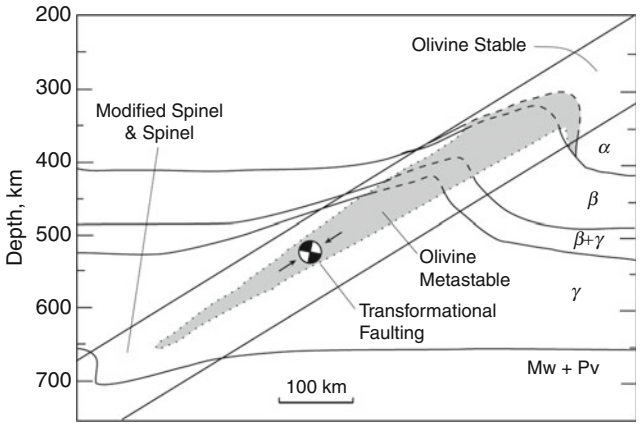


Seismicity, Subduction Zone, Figure 5 Sketch of a model explaining the stress state within the subducting slab where gravitational forces act on excess mass within the slab (Isacks and Molnar, 1971, copyright by the American Geophysical Union).

than in the ambient mantle. Moreover, the olivine to β-spinel transition must be kinetically hindered within the cold slab, since the relatively colder temperatures in the slab inhibit reaction rates. A wedge-shaped zone of

metastable peridotite thus formed in the cold core of the slab persists to the depth shown by the shaded area in Figure 6 (Kirby, 1995).

Upward deflections of the olivine to β-spinel transition at 410 km and the β-spinel to γ-spinel transition at 550 km yield negative buoyancy forces by stabilizing the dense higher pressure phases in the slab surrounded by the less dense lower pressure phases in the ambient mantle. On the contrary, the downward deflection of the γ-spinel to perovskite + magnesiowustite transition at about 660 km yields positive buoyancy forces. Moreover, the metastable olivine wedge yields positive buoyancy forces. Downdip compression at depths deeper than about 300 km, observed for all the subduction zones (Figure 4), can be well explained by the compressional stress generated by the combination of the buoyancy forces produced by the metastable olivine wedge, the downward deflection of the perovskite-forming transition, and the force resisting slab penetration into the lower mantle (Chen et al., 2004). It is expected that the downdip compressional stress caused by the metastable olivine wedge is smaller for warm (young) and slow slabs. Chen et al. (2004) found a good correlation between the thermal parameter of the slab and the nature of the downdip stress at depths shallower than about 300 km in accordance with this expectation: warm, slow slabs tend to have tensional stress; and cold, fast slabs tend to have compressional stress.

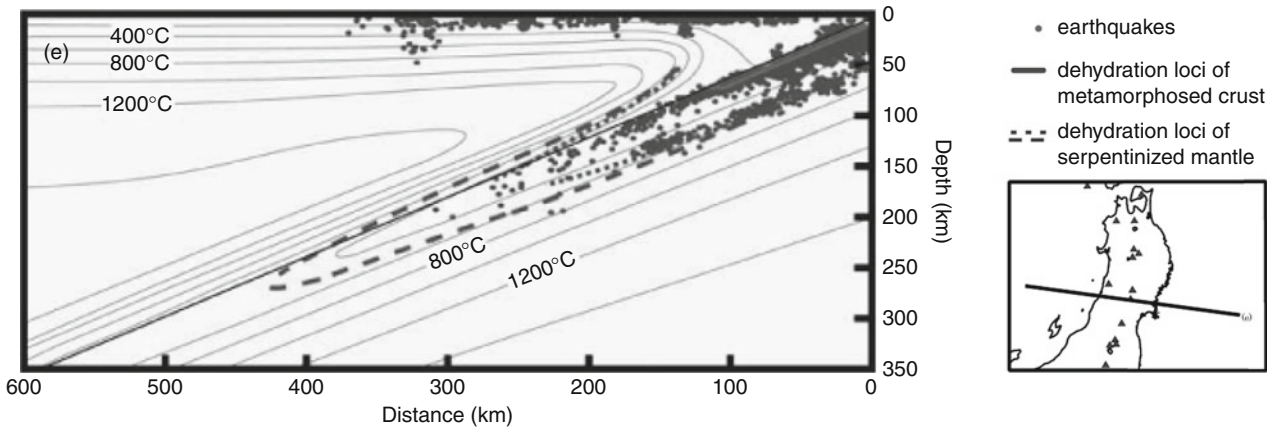


Seismicity, Subduction Zone, Figure 6 Schematic diagram of the mineralogy within a subducting slab (Kirby, 1995, copyright by the American Geophysical Union).

Intermediate-depth earthquakes: double seismic zone

It is known that intermediate-depth intraslab earthquakes form a double seismic zone in several subduction zones. This was first clearly found beneath northern Honshu, in northeast Japan (Figure 7), where the deep seismic zone is composed of two planes at depths of 70–150 km that are parallel to each other and separated by about 30 km (Hasegawa et al., 1978). Upper-plane earthquakes, most of which occur in the crust of the slab, have DC-type focal mechanisms, while lower-plane ones, occurring in the mantle of the slab, are characterized as DE-type earthquakes. Engdahl and Scholz (1977) explained the formation of such a double-planed deep seismic zone by the unbending of the subducting oceanic plate at this depth range. Later studies showed, however, that DC-type upper-plane and DE-type lower-plane events are not always seen and that in some subduction zones both the upper and lower planes have DC stress. These observations indicate that the unbending force may not be the main cause of the formation of the double seismic zone, although it must be acting on the plate.

Similar to ordinary shallow earthquakes, most intermediate-depth and deep earthquakes have double-couple radiation patterns and little or no isotropic



Seismicity, Subduction Zone, Figure 7 Double seismic zone and dehydration loci of oceanic crust and mantle for central Tohoku, northeast Japan, projected on an across-arc vertical cross section (Yamasaki and Seno, 2003, copyright by the American Geophysical Union). The solid line and broken and dotted lines show the dehydration loci of the crust and mantle, respectively. Dots denote earthquake hypocenters. Medium thick and thin lines are the upper plate interface and isotherm contours, respectively.

component (Frohlich, 2006). However, it has been thought that intermediate-depth and deep earthquakes are mechanically different from shallow earthquakes that are generated by brittle fracture. This is because normal stress on the fault becomes too high to cause ordinary brittle shear faulting. At depths of 60–650 km, the overburden produces confining pressures of 2–23 GPa, yielding the strength of the fault of about the same order of magnitude. Earthquake occurrence requires a shear stress exceeding that strength, but it is unlikely that such a high shear stress is actually produced at such depths in the slab. Even if some process in the slab can produce such a high shear stress, it cannot be maintained because ductile flow will occur. Therefore, some special mechanism is necessary to generate earthquakes at such great depths.

One possible mechanism that decreases the strength of the fault for intermediate-depth earthquakes is dehydration embrittlement (Raleigh and Paterson, 1965; Green and Houston, 1995). Before its subduction, the oceanic plate contains a considerable amount of water as hydrated minerals. The temperature and pressure of the plate rise as it subducts and decompose the hydrated minerals and discharge water. The water reduces the effective normal stress and enables brittle fracture. Dehydration embrittlement also seems to explain the formation of the double seismic zone as discussed below.

If dehydration embrittlement is the mechanism generating intermediate-depth intraslab earthquakes, intraslab earthquakes should not be expected to occur everywhere in the slab, but only where hydrated minerals exist, particularly at facies boundaries where the water content changes. It has become increasingly clear that intermediate-depth intraslab earthquakes do seem to occur in such areas.

In order to examine whether or not dehydration embrittlement is the cause of intraslab earthquakes, we need to know the location within the slab where the hydrated minerals are decomposed and the water is discharged. Yamasaki and Seno (2003) determined the locations in the slab where the dehydration decompositions of serpentinized slab mantle and metamorphosed slab crust occur based on experimentally obtained phase diagrams of rocks. They compared these locations with the spatial distribution of intraslab earthquakes. The result for the Pacific slab under northeast Japan is shown in Figure 7. The estimated dehydration loci of the crust and mantle roughly correspond to the location of the upper and lower planes of the double seismic zone. They obtained the same results for the other five subduction zones. This explains well why double seismic zones are formed and why earthquakes in the lower plane have a planar distribution near the center of the slab mantle.

The above results support the dehydration embrittlement model of intraslab earthquakes. However, whether hydrated minerals actually exist at depths where lower-plane earthquakes occur (the deeper part of the mantle that lies at a maximum of about 40 km below the upper slab boundary) is still a point of contention. In this regard, further validation is necessary for the generation of lower-plane events.

Formation of a belt of intraslab seismicity in the slab crust beneath northeast Japan

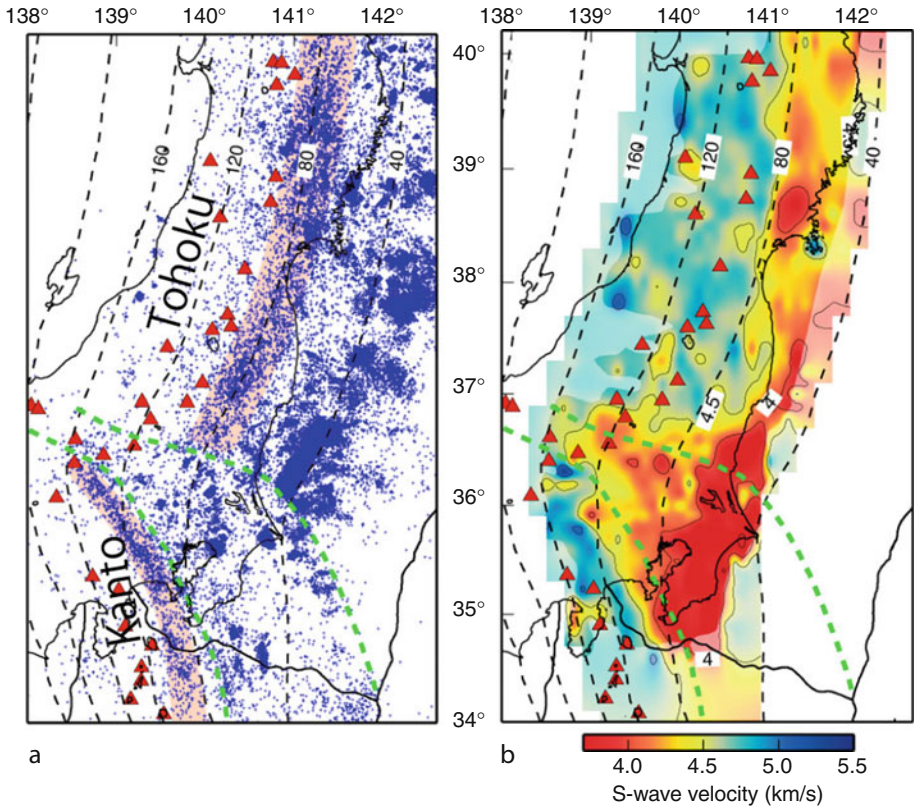
Recent studies have shown a more detailed spatial distribution of upper-plane earthquakes and revealed its clear relationship with the slab structure. This, as described below, can be explained well if the cause of intraslab earthquakes is dehydration embrittlement.

As indicated by the pale-pinked zones in Figure 8a, earthquakes in the upper plane of the double seismic zone beneath northeast Japan form a remarkable belt-like seismic zone at depths between 70 and 90 km, parallel to iso-depth contours of the upper surface of the Pacific slab. The across-arc vertical cross section of earthquakes in Figure 9 also shows that there is a concentration of earthquakes at depths of 70–90 km, which corresponds to the belt-like seismicity indicated in Figure 8a. If the cause of intraslab earthquakes is dehydration embrittlement, this concentration of seismicity is expected to correspond to the location in the slab crust where hydrated minerals are present, particularly around the facies boundary where dehydration occurs.

This belt-like seismicity indicated by the pale-pinked zones is not parallel to iso-depth contours, but is oblique to them and deepens locally in Kanto (Figure 8a). This local deepening is thought to be caused by the shield effect of the Philippine Sea slab, which is located immediately above the Pacific slab. The two slabs are in direct contact under Kanto, and the contact zone between the two is indicated as an area enclosed by two broken lines in the figure. The downdip end of the contact zone is parallel to the belt-like seismicity beneath Kanto, which is oblique to the iso-depth contours of the upper plate interface. Such a clear correspondence indicates that under Kanto, the Philippine Sea slab prevents the Pacific slab from being heated up by the mantle wedge and delays the phase transformation in the slab crust, resulting in the locally deepened belt-like seismicity in that area (Hasegawa et al., 2007).

Phase transformation of the slab crust causes an increase in seismic wave velocity. Therefore, we can verify from seismic tomography whether the phase transformation is the cause of the belt-like seismicity. Distribution of S-wave velocity in the slab crust (Figure 8b) shows that the low-velocity layer persists down to about 80 km depth in Tohoku; deeper than that, S-wave velocity becomes high. On the other hand, in Kanto, the depth range of the low-velocity layer deepens locally in the zone of contact with the Philippine Sea slab. This means the distribution of S-wave velocity also confirms the prediction that the phase transformation is delayed.

Direct comparison with mineralogical research results does not allow us to verify whether the phase transformation accompanying dehydration actually occurs at the depth of this belt-like seismicity. Though accurate



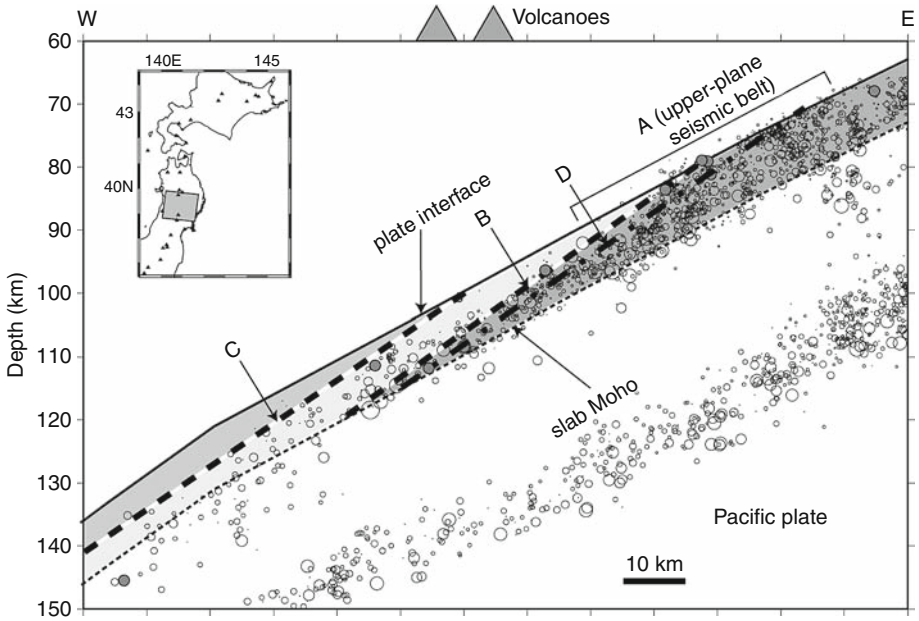
Seismicity, Subduction Zone, Figure 8 (**a**) Earthquake epicenter distribution and (**b**) S-wave velocity distribution in the subducting Pacific slab crust beneath northeast Japan (Nakajima et al., 2009, copyright by the American Geophysical Union). Pale-pinked zones in (**a**) show the belt-like seismic zone. S-wave velocity perturbations on a curved plane 5 km below the upper surface of the slab are shown in (**b**) by the color scale at the bottom. Iso-depth contours of the upper surface of the slab are shown by broken curves. The zone of contact with the overlying Philippine Sea slab is enclosed by green broken lines.

temperature distribution in the slab is necessary for such a detailed examination, unfortunately, many uncertainties remain regarding the estimation of the temperature in the slab. However, it is possible to make at least a rough estimate. In the vertical cross section in Figure 9, the facies boundaries estimated based on the phase diagram of midoceanic ridge basalt (MORB) by Hacker et al. (2003) and the geotherm by Peacock and Wang (1999) are shown as broken lines. The concentration of earthquakes forming the belt-like seismicity is found around the facies boundary in the shallower areas (B in the figure). That approximately agrees with the prediction based on the dehydration embrittlement model. As another example, the facies boundary determined by using the MORB phase diagram by Omori et al. (2009) and the subducted slab geotherm by van Keken et al. (2002) are indicated as a dashed-dotted line (D) in the figure. In this case also, the facies boundary approximately agrees with the location of the belt-like seismicity. Since many uncertainties remain in the estimation of the slab temperature, we need further validation for the location of the facies boundary. At any rate, the facies boundary must be strongly dependent not on pressure but on temperature.

Deep earthquakes

As described in previous sections, seismic observations seem to support the dehydration embrittlement hypothesis for the generation of intermediate-depth earthquakes, at least those in the crust of the slab. Focal depth distribution of global seismicity (Figure 3) shows that the frequency of earthquakes, after a nearly monotonic decrease from shallow depths, recovers at a depth of about 300 km and gradually increases until the local maximum at 500–600 km. This may suggest that a different mechanism is working for the generation of deep earthquakes.

Several hypotheses have been proposed for the generation of deep earthquakes. There exists the difficulty in generating deep earthquakes as already described in section "Intermediate-depth earthquakes: double seismic zone." Three mechanisms out of those proposed so far seem to overcome the difficulty. One is the shear heating



Seismicity, Subduction Zone, Figure 9 Across-arc vertical cross section of intraslab earthquakes in central Tohoku, northeast Japan (Kita et al., 2006, copyright by the American Geophysical Union). Earthquakes are shown by open circles. The estimated jadeite lawsonite blueschist–lawsonite amphibole eclogite facies boundary (B) and the lawsonite amphibole eclogite–eclogite facies boundary (C) of the MORB (Hacker et al., 2003) are denoted by thick broken lines. The facies boundary (D) based on Omori et al. (2009) is also shown by a dashed-dotted line.

instability of creep (Griggs and Handin, 1960; Ogawa, 1987). Plastic creep of minerals produces heat, while high temperatures promote creep. Therefore, positive feedback promoting creep would arise if the creep of a thin layer occurs at a higher rate than the heat produced by the creep diffuses into the surrounding area. Thus, the creep accelerates, and finally the thin layer of material begins to melt, producing dynamic slip on the layer, that is, occurrence of an earthquake.

The second hypothesis is the transformational faulting (Kirby et al., 1987; Green and Burnley, 1989). Olivine in the mantle of the slab transforms to β-spinel associated with the subduction of the slab. The phase transformation does not occur simultaneously but originates sporadically here and there, forming small lens-shaped spinel phases. The phase transformation proceeds as these spinel lenses expand. The spinel lenses will be formed in a direction perpendicular to the maximum compressive stress axis due to volume reduction occurring in the phase transformation. At the same time, compressive stress will arise at the tip of each spinel lens due to volume reduction as the phase transformation proceeds. This compressive stress promotes the phase transformation. In this way, the spinel lens expands as if it were a crack. Green and Burnley called this spinel lens anticrack. At some critical anticrack density, those anticracks that are aligned on a plane directed obliquely to the maximum compressive axis at an acute angle are connected with each other and become unified, producing a dynamic slip on the plane, that is, the occurrence of an earthquake.

The last one is the dehydration embrittlement. Although this hypothesis seems to work for the generation of intermediate-depth earthquakes as already explained, some researchers argue that the dehydration reaction continues to occur in the mantle of the slab at greater depths and that dehydration embrittlement is responsible also for deep earthquakes (Omori et al., 2004).

It is not clear which of the three mechanisms is actually working, or if another completely different mechanism is responsible for generating deep earthquakes. More investigations are required for an exact understanding of the mechanism by which deep earthquakes are generated.

Summary

Recent investigations have significantly developed our understanding of generation mechanism for earthquakes in subduction zones. Asperity model seems to be applicable to interplate earthquakes. This gives us a theoretical background for long-term earthquake forecast: Place and magnitude of future earthquakes can be estimated from information on locations and sizes of asperities on the plate boundary. Dehydration embrittlement model seems to work as the generation mechanism for intraslab intermediate-depth earthquakes, particularly for earthquakes in the slab crust. Aqueous fluids expelled by the

dehydration reaction in the subducted plate migrate upward, finally reaching the arc crust. Recent investigations have shown that the fluids thus supplied originally from the subducting plate play an important role in generating shallow inland intraplate earthquakes. Generation mechanism for deep earthquakes is still an open question. More systematic and intensified studies are expected to resolve it.

Bibliography

Bilek, S. L., and Lay, T., 2002. Tsunami earthquakes possibly widespread manifestations of frictional conditional stability. *Geophysical Research Letters*, **29**(14), 1673, doi:10.1029/2002GL015215.

Bina, C. R., Stein, S., Marton, F. C., and Van Ark, E. M., 2001. Implications of slab mineralogy for subduction dynamics. *Physics of the Earth and Planetary Interiors*, **127**, 51–66.

Chen, Po-Fei, Bina, C. R., and Okal, E. A., 2004. A global survey of stress orientations in subducting slabs as revealed by intermediate-depth earthquakes. *Geophysical Journal International*, **159**, 721–733.

Dixon, T. H., and Moore, J. C., 2007. *The Seismogenic Zone of Subduction Thrust Faults, Margins Theoretical Institute and Experimental Earth Science Series*. New York: Columbia University Press.

Engdahl, E. R., and Scholz, C. H., 1977. A double Benioff zone beneath the central Aleutians: an unbending of the lithosphere. *Geophysical Research Letters*, **4**, 473–476.

Engdahl, E. R., van der Hilst, R., and Buland, R., 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America*, **88**, 722–743.

Frohlich, C., 2006. *Deep Earthquakes*. New York: Cambridge University Press.

Green, H. W., and Burnley, P. C., 1989. A new self-organizing mechanism for deep-focus earthquakes. *Nature*, **341**, 733–737, doi:10.1038/341733a0.

Green, H. W., and Houston, H., 1995. The mechanics of deep earthquakes. *Annual Review of Earth and Planetary Sciences*, **23**, 169–213.

Griggs, D. T., and Handin, J., 1960. Observations on fracture and a hypothesis of earthquakes. In Griggs, D. T., and Handin, J. (eds.), *Rock Deformation: Geological Society of America. Memoirs, 79*. New York: Geological Society of America, pp. 347–373.

Hacker, B. R., Peacock, S. M., Abers, G. A., and Holloway, S. D., 2003. Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *Journal of Geophysical Research*, **108**(B1), 2030, doi:10.1029/2001JB001129.

Hasegawa, A., Umino, N., and Takagi, A., 1978. Double-planed structure of the deep seismic zone in the northeastern Japan arc. *Tectonophysics*, **47**, 43–58.

Hasegawa, A., Nakajima, J., Umino, N., and Miura, S., 2005. Deep structure of the northeastern Japan arc and its implications for crustal deformationand shallow seismic activity. *Tectonophysics*, **403**(1–4), 59–75.

Hasegawa, A., Nakajima, J., Kita, S., Okada, T., Matsuzawa, T., and Kirby, S., 2007. Anomalous deepening of a belt of intraslab earthquakes in the pacific slab crust under Kanto, central Japan: possible anomalous thermal shielding, dehydration reactions, and seismicity caused by shallower cold slab material. *Geophysical Research Letters*, **34**, L09305, doi:10.1029/2007GL029616.

Hyndman, R. D., Wang, K., et al., 1995. Thermal constraints on the seismogenic portion of the southwestern Japan subduction thrust. *Journal of Geophysical Research*, **100**, 15373–15392.

Iio, Y., Sagiya, T., and Kobayashi, Y., 2004. Origin of the concentrated deformation zone in the Japanese Islands and stress accumulation process of intraplate earthquakes. *Earth Plants Space*, **56**, 831–842.

Isacks, B. L., and Molnar, P., 1971. Distribution of stresses in the descending lithosphere from a global survey of focal mechanism solutions of mantle earthquakes. *Reviews of Geophysics and Space Physics*, **9**, 103–174.

Kanamori, H., 1977. Seismic and aseismic slip along subduction zones and their tectonic implications. In Ewing, M. (ed.), *Island Arcs, Deep Sea Trenches and Back-Arc Basins*. Washington, DC: American Geophysical Union, Vol. I, pp. 163–174.

Kirby, S., 1987. Localized polymorphic phase transformations in high-pressure faults and applications to the physical mechanism of deep earthquakes. *Journal of Geophysical Research*, **92**(B13), 789–800, doi:10.1029/JB092iB13p13789.

Kirby, S., 1995. Intraslab earthquakes and phase changes in subducting lithosphere. *Reviews of Geophysics*, **33**(S1), 287–297.

Kita, S., Okada, T., Nakajima, J., Matsuzawa, T., and Hasegawa, A., 2006. Existence of a seismic belt in the upper plane of the double seismic zone extending in the along-arc direction at depths of 70–100 km beneath NE Japan. *Geophysical Research Letters*, **33**, L24310, doi:10.1029/2006GL028239.

Lay, T., and Kanamori, H., 1981. Fundamental studies, laboratory investigations and models. In Ewing, M., Simpson, D. W., and Richards, P. G. (eds.), *An International Review, Earthquake Prediction*. Washington, DC: AGU, pp. 579–592.

Miura, M., Sato, T., Hasegawa, A., Suwa, Y., Tachibana, K., and Yui, S., 2004. Strain concentration zone along the volcanic front derived by GPS observations in NE Japan arc. *Earth Planets Space*, **56**, 1347–355.

Nakajima, J., and Hasegawa, A., 2007. Deep crustal structure along the Niigata-Kobe tectonic zone, Japan: its origin and segmentation. *Earth Planets and Space*, **59**, e5–e8.

Nakajima, J., Matsuzawa, T., Hasegawa, A., and Zhao, D., 2001. Three-dimensional structure of Vp, Vs, and Vp/Vs beneath northeastern Japan: implications for arc magmatism and fluids. *Journal of Geophysical Research*, **106**, 843–857.

Nakajima, J., Tsuji, Y., and Hasegawa, A., 2009. Seismic evidence for thermally-controlled dehydration reaction in subducting oceanic crust. *Geophysical Research Letters*, **36**, L03303, doi:10.1029/2008GL036865.

Ogawa, M., 1987. Shear instability in a viscoelastic material as the cause of deep focus earthquakes. *Journal of Geophysical Research*, **92**, 13801–13810.

Oleskevich, D. A., Hyndman, R. D., and Wang, K., 1999. The updip and downdip limits to great subduction earthquakes: thermal and structural models of Cascadia, south Alaska, SW Japan, and Chile. *Journal of Geophysical Research*, **104**, 14965–14991.

Omori, S., Komabayashi, T., and Maruyama, S., 2004. Dehydration and earthquakes in the subducting slab: empirical link in intermediate and deep seismic zones. *Physics of the Earth and Planetary Interiors*, **146**, 297–311.

Omori, S., Kita, S., Maruyama, S., and Santosh, M., 2009. Pressure – temperature conditions of ongoing regional metamorphism beneath the Japanese islands. *Gondwana Research*, **16**, 458–469.

Peacock, S. M., and Wang, K., 1999. Seismic consequences of warm versus cool subduction metamorphism: examples from southwest and northeast Japan. *Science*, **286**, 937–939.

Raleigh, C. B., and Paterson, M. S., 1965. Experimental deformation of serpentinite and its tectonic implications. *Journal of Geophysical Research*, **70**, 3965–3985.

Ruff, L. J., and Kanamori, H., 1980. Seismicity and the subduction process. *Physics of the Earth and Planetary Interiors*, **23**, 240–252.

Sagiya, T., Miyazaki, S., and Tada, T., 2000. Continuous GPS array and presentday crustal deformation of Japan. *Pure and Applied Geophysics*, **157**, 2303–2322.
Scholz, C. H., 1998. Earthquakes and friction laws. *Nature*, **391**, 37–42.
van Keken, P. E., Kiefer, B., and Peacock, S. M., 2002. High-resolution models of subduction zones: implications for mineral dehydration reactions and the transport of water into the deep mantle. *Geochemistry, Geophysics, Geosystems*, **3**, 10, doi:10.1029/2001GC000256.
Yamasaki, T., and Seno, T., 2003. Double seismic zone and dehydration embrittlement of the subducting slab. *Journal of Geophysical Research*, **108**(B4), ESE9–1, doi:10.1029/2002JB001918.

Cross-references

Earthquake, Focal Mechanism
Great Earthquakes
Seismicity, Intraplate

SEISMOGRAM INTERPRETATION

Ota Kulhanek[1], Leif Persson[2]
[1]Department of Earth Sciences, Section of Seismology, Uppsala University, Uppsala, Sweden
[2]Department of Underwater Research, FOI, Stockholm, Sweden

Synonyms

Anatomy of seismograms; Reading seismic records; Structure of seismograms

Definition

Seismogram interpretation. Identification of various phases appearing on seismograms (measurable changes in frequency and/or amplitude), including the recognition of the waves with respect to possible travel paths through the Earth.

Introduction

A seismogram is the recording of earth motion (displacement, velocity, or acceleration) as a function of time. Its appearance reflects the combined effects of the source, the propagation path, the characteristics of the recording instrument (see *Seismic Instrumentation*), and the ambient noise due to the specific conditions at the particular recording site. To decipher the often complicated form of seismogram traces requires an extensive knowledge of source physics, structure of the Earth, wave propagation, and earthquake geography. In the early years of observational seismology, say, prior to 1960, the analyst was essentially reading short- and long-period analog seismograms made on smoked, photographic, or heat-sensitive paper or by ink-pen recorders. Approximately, in the middle of the 1980s, the situation changed dramatically. The advent and extensive deployment of broadband digital seismometers with large dynamic ranges complemented with an access to fast, powerful computers opened for seismogram interpreters new possibilities never expected before. Today, the interpreter performs easily a set of operations that significantly improve and accelerate the process of phase identification on digital seismic records. These include, for example, rotation of seismogram components and particle motion techniques. Digital multichannel data serve as input for frequency-wave number analysis and polarization filtering. Synthetic seismograms model observations and phase onsets revealed on seismograms can be associated within narrow time windows with theoretically predicted onset arrival times.

Velocity models for radially stratified Earth have been developed to provide travel times for major seismic waves for the purpose of earthquake location and phase identification. In 1987, IASPEI initiated an effort to update the JB Tables (Jeffreys and Bullen, 1940), which have served for 50 years, by constructing new global travel-time tables. With access to an extensive ISC data set (1964–1987) this effort resulted in two velocity models: *iasp91* (Kennett and Engdahl, 1991) and *sp6* (Morelli and Dziewonski, 1993) deduced from empirical smoothed travel-time curves of the main phases. The two models reveal only small differences in predicted travel times. With respect to the older JB Tables, the most significant differences were found in the upper mantle and core (see *Earth's Structure, Global*). New, *iasp91*, travel-time tables, which include also the ancillary phases, were calculated from the velocity model (Kennett, 1991). For teleseismic *P* waves, the new tables are in average 1.8–1.9 s faster than the JB Tables. The *iasp91* model was slightly modified in 1995 and a new model *ak135* was proposed (Kennett et al., 1995). The model is suited for predicting the arrival times of a wide variety of seismic phases. It is routinely used by major agencies (ISC, NEIC) in event locations and phase identifications.

At first sight, it seems that there is a contradiction between the duration of the rupture at the source, which takes between a fraction of a second and a few minutes (depending on magnitude), and the length of the observed seismogram, which for large and distant shocks can extend over several hours. The length of the seismogram depends primarily on various wave propagation effects such as reflection, refraction, mode conversion, dispersion, etc., and has little to do with the duration of the quake. A seismogram, especially from a distant earthquake, will often show a number of more or less distinct waves, commonly called phases, distributed in time, which have traveled along different propagation paths and, which were subjected to different mode conversions. Numerous manuals for seismogram interpretation have been available since the early 1950s. Some of them have been of local/regional importance, but some have gained worldwide recognition. The latter include works of Neumann (1951), Simon (1968), Willmore (1979), Payo (1986), and Kulhanek (1990). The most recent is the manual edited by Borman (2002).

Essentially, there are two types of seismic waves, *P* and *S* body waves (see *Body Waves*) and Love (*LQ*) and Rayleigh (*LR*) surface waves (see *Surface Waves*). The most important characteristics of body and surface waves, which are invaluable in any seismogram interpretation, are the following:

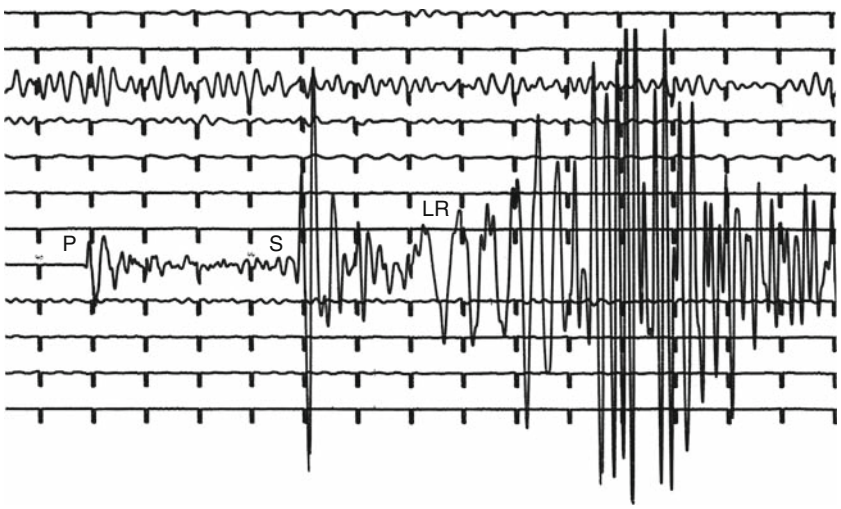
- Different waves travel with different velocities. At any epicentral distance, *P* is recorded first, followed by *S*, *LQ*, and *LR*.
- Body waves produce records in the high-frequency range, from 0.1 to 10 Hz, while surface waves occupy low frequencies of about 0.1 Hz and lower. Exceptions will be described later.
- Due to different polarization, *P*-waves are usually better displayed on vertical-component seismograms while *S*-waves are often best on horizontal-component records. Fluids do not sustain shear strain and therefore *S*-waves do not travel through liquid parts of Earth's interior. *LQ*-waves are exhibited only on horizontal-component records, while *LR*-waves are received by both the vertical and horizontal seismometers. For shallow shocks, surface waves usually dominate the seismogram.
- Surface waves exhibit an important property called velocity dispersion (see *Surface Waves*). In practice, this means that the long period surface waves approach the station first and are recorded ahead of the "slower" shorter waves. Ideally, the seismogram of *LQ*- or *LR*-waves will start with long-period motion, which gradually, as time increases, will turn into shorter and shorter periods.

Some of these features are demonstrated on the seismogram displayed in Figure 1. There is a sharp *P* onset followed after approximately 4 min by a clear *S*. About 2 min after *S* we observe the arriving *LR* (vertical-component seismogram). The clearly dispersive character of the recorded *LR* in this case manifests a continental propagation path signature. The large *LR* amplitudes indicate a shallow-focus depth. Except for the first *P* onset, all later arrivals are contaminated by codas of preceding phases so that, on the records, there is virtually no interval of quiescence between individual phases.

In the following description, we introduce several categories of seismic events. The classification is based on the distance between the event and the recording site, that is, on the epicentral distance, which in turn governs propagation paths along which seismic waves travel through the Earth's interior. Events recorded at distances shorter than about 1° will be called local events. Regional earthquakes are shocks recorded at distances between approximately 1° and 10°. For epicentral distances 0–10°, seismic waves propagate through the crust and/or along Moho. Accordingly, we call these waves, crustal waves. At recording distances 10–103°, the waves travel mainly through the mantle. For distances larger than about 103°, seismic waves enter the core or are diffracted along the core-mantle boundary (so-called core waves). Events observed at distances larger than 20° (or 30°) are collectively called teleseisms.

Crustal waves; recording distances 0–10°

In a one-layer continental crust, a source located within the crust radiates both *P* and *S* waves, which will be recorded along the Earth's surface. These waves are encoded as *Pg* and *Sg*. The subscript *g* refers to granitic layer. In continental earthquakes it has been often observed that *Sg* has the largest amplitude. The nomenclature of seismic phases used in the present article is that of the IASPEI Standard (Storchak et al., 2003 or *Seismic*

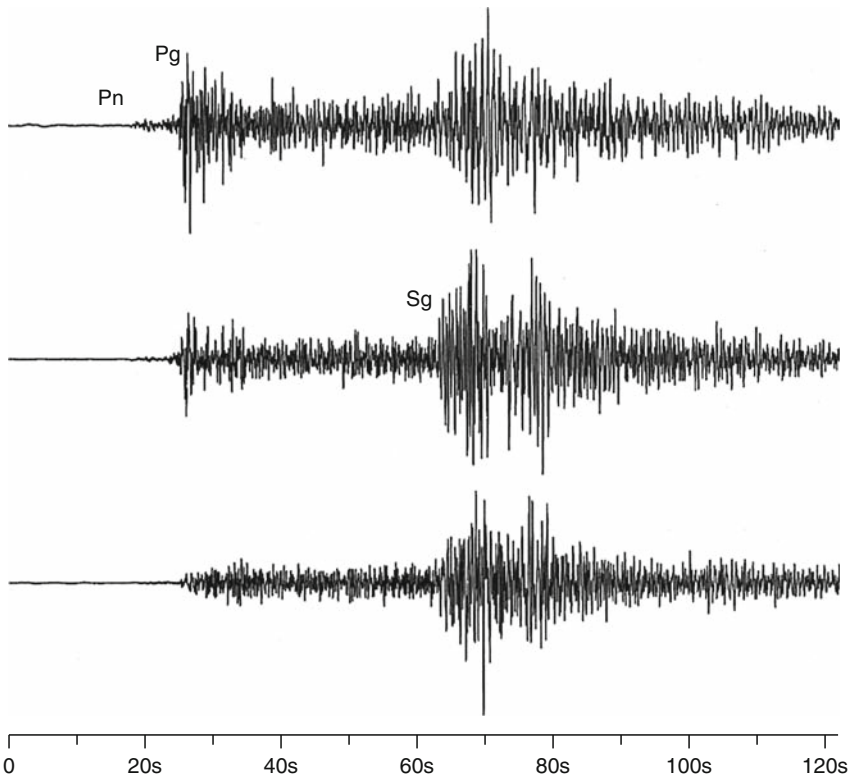


Seismogram Interpretation, Figure 1 Long-period, vertical-component, ink-pen seismogram of the Greek earthquake of February 25, 1981 (M = 6.3, h = 33 km) made at UPP (Uppsala), Sweden, at an epicentral distance of 22°. There is 1 min between successive time marks (small downward offsets).

Phase Names: IASPEI Standard). A reflected wave is also possible from the outer side of the Moho, and the corresponding reflected *P* and *S* waves are labeled as *PmP* and *SmS*, respectively. At larger distances (beyond the critical distance), we shall record so-called head waves, that is, *P* and *S* waves refracted along the Moho discontinuity and labeled *Pn* and *Sn*. Head waves propagate with velocities of the uppermost mantle. Note that *Pg* and *Sg* exist for all epicentral distances from Δ=0 and outwards, while *Pn* and *Sn* cannot be observed at distances shorter than the critical distance, which for a continental crust is about 100 km. Waves reflected from Moho are rather scarce and difficult to identify. The best chance to discern *PmP* or *SmS* on records is at very short epicentral distances where the contamination by *Pg* and *Pn* (or *Sg* and *Sn*) is not severe.

At epicentral distances shorter than about 10°, records are complicated because of lateral heterogeneities in the crust. In a two-layer crustal model, which includes the Conrad discontinuity, we record new phases. For epicentral distances larger than about 100 km, these are the refracted *P* and *S*, traveling along the Conrad discontinuity. An asterisk in the superscript position, P^* and S^*, indicates this phase. An alternative code is *Pb* and *Sb*. The subscript *b* refers to basaltic layer.

Let us assume a continental crust with propagation velocities for *Pg*, P^*, and *Pn* of 6.0, 6.6, and 8.0 km/s, respectively. At short distances, less than about 150 km or so, the first seismic wave arriving at the recording station is *Pg*. For distances larger than critical but less than about 150 km/s, *Pg* is followed by P^* and *Pn*, in this order. P^* and *Pn* travel with velocities significantly higher than *Pg*. Therefore, at distances larger than approximately 150–200 km crustal waves change their order of arrival. For distances larger than about 200 km, the first arriving phase is *Pn*, next arrives P^*, and then *Pg*. This is true only for continental travel paths. Seismograms from earthquakes beneath the sea bottom, made at islands or coastal stations, will not show *Pg* or *Sg* since there is no granitic layer in the oceanic crust. Similarly, quakes originating in the lower crust, beneath the Conrad discontinuity, do not produce *Pg* or *Sg* phases. Hence, first arrivals on records from these events will be *Pn* or P^*. With a certain time delay following the *P* phases, proportional to the epicentral distance, the crustal *S* waves arrive in the same order as *P* waves. For local events the order of *S* onsets will be *Sg*, S^*, *Sn* while for events from distances larger than about 300 km, we observe first *Sn* followed by S^* and *Sg*. As an example, a record from a regional earthquake is displayed in Figure 2. The first discernible phase



Seismogram Interpretation, Figure 2 The Kaliningrad, Russia, earthquake of September 21, 2004 (M = 5.0, h = 15 km) recorded at broadband station GOT (Gotland), Sweden, at an epicentral distance of 335 km. From top to bottom: vertical, N–S and E–W components. There are 10 s between successive time marks.

is *Pn*, weakly recorded on the vertical and N-S channels. It is followed after about 6 s by strong *Pg*, and after another 39 s by clear *Sg*, visible on all three components. *Sn* expected to precede *Sg* by 9 s cannot be identified on the seismograms.

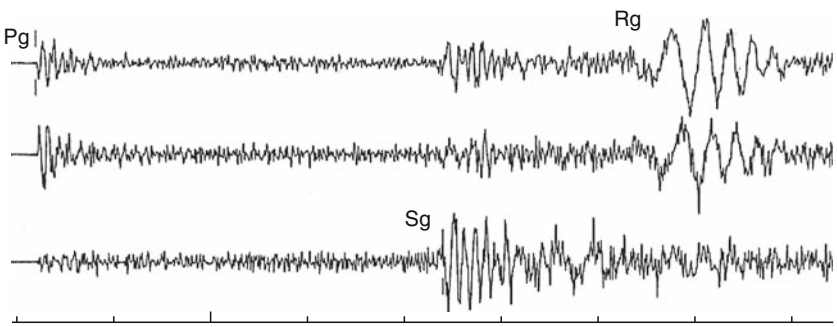
Short-period *S* waves, multiply reflected between the free surface and Moho or other crustal velocity discontinuities, interfere with each other and give rise to a wave group labeled *Lg*. There is no clear distinction between *Sg* and *Lg*. *Lg* waves propagate as guided waves, supercritically incident on the Moho and multiply reflected within the (continental) crust with a velocity of about 3.5 km/s. They may dominate the seismograms made usually at a distance of 5° and larger. *Lg* are best seen on horizontal-component records. Some rare cases are known where *Lg* propagated over distances of several thousand kilometers (Kulhanek, 1990).

Near-surface local and regional events (earthquakes, industrial explosions, rockbursts, etc.) also generate short-period surface wave of Rayleigh type, labeled *Rg*. The presence of *Rg* in the seismogram is a reliable indicator of a very shallow event with focal depth of the order of one or a few kilometers. On the other hand, if *Rg* waves are absent (epicentral distance of several hundred kilometers or less), we are concerned with a deeper natural event, that is, with a crustal quake at a depth most likely between 5 and 25 km, since all types of man-made events can be excluded. *Rg* waves travel with a velocity of 3 km/s or slightly higher and are usually well recorded up to distances of about 600 km (Båth, 1983). These waves are best, but not exclusively, seen on vertical-component records. An example of a recorded *Rg* wave is displayed in Figure 3. The *Rg* phase is well developed (large amplitudes, clear dispersion) indicating a focal depth of 2 km or less.

Island and coastal seismographic stations frequently record so-called *T* waves (tertiary waves) arriving after *P* and *S* (see *T Waves*). They are characterized by propagation within the oceans as ordinary sound waves emitted by earthquakes near the sea bottom or by submarine volcanic eruptions. The propagation of *T* waves is very efficient and observations at distances as large as about 80° have been reported. Observations of *T* phases have proved useful in discriminating between underground nuclear explosions, detonated beneath oceanic islands and tectonic earthquakes (Adams, 1979). On records, *T* phases often exhibit rather monochromatic oscillations (periods usually less than 1 s) with a gradual increase and decrease of amplitudes of total duration up to several minutes. There is no sharp onset in the *T* phase group, which creates difficulties when reading the *T* arrival times. In general, there is a great variety in the appearance of *T* phases due to the dependence upon the bottom topography in the vicinity of generation, oceanic stratification, and water-land conversion and transmission.

Appearance of seismograms made at local and/or regional distances varies from place to place mainly due to lateral variations in crustal structure. It is, therefore, difficult to list generally valid clues for record interpretation. Nevertheless, following principles may guide the analyst to read seismograms and evaluate software solutions of local and regional earthquakes.

- Predominant periods of recorded crustal waves *Pg*, P^*, *Pn*, *Sg*, S^*, *Sn*, etc., are normally less than 1 s. *Rg* periods are usually not longer than several seconds.
- Often, *Sg* has the largest amplitude (when large short-period *Rg* is missing), best on horizontal-component records. At larger distances, *Lg* may dominate the records.
- For epicentral distances less than about 200 km (depending upon the crustal structure and focal depth), the first arriving phase is *Pg*. For larger distances, *Pn* arrives first.
- Near-surface events, from distances less than about 600 km, often generate short-period *Rg* with clear dispersion, best seen on vertical channels.
- Local and regional events of low or moderate magnitude are characterized by short total record duration, usually not exceeding several minutes.
- Island and coastal seismographic stations frequently record various types of *T*-phases.



Seismogram Interpretation, Figure 3 Weak, near-surface event ($M \sim 2$) in southern Sweden, recorded at broadband station NYN (Nynäshamn) at an epicentral distance of 36 km. From top to bottom: vertical, N–S and E–W components. There are strong onsets of *Pg* (vertical, N–S) and *Sg* (E–W) and a clear *Rg*-phase with distinct dispersion, about 2 s after *Sg*, best developed on the vertical and N–S components. There is 1 s between successive time marks at the bottom. (Note: epicenter almost due north of NYN).

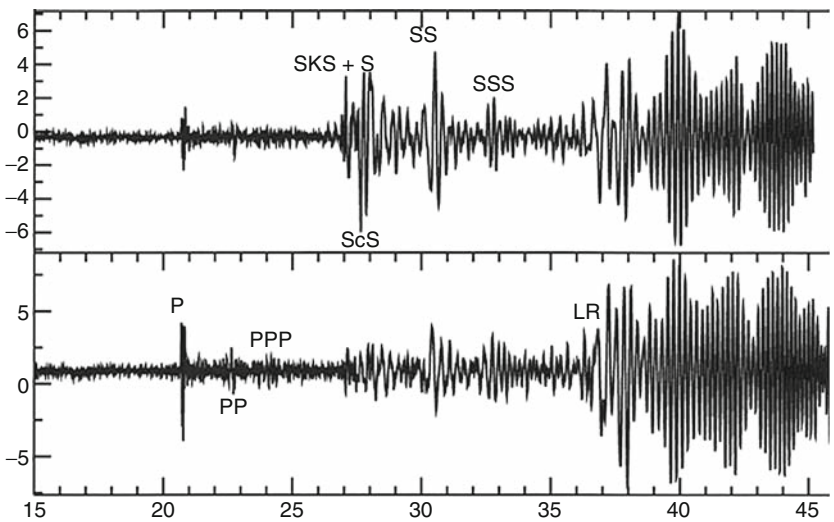
Body waves traversing the mantle; recording distances 10–103°

Mantle, in contrast to the overlying crust, may, in the first approximation, be considered as laterally homogeneous. At distances around 10°, *Pn* and *Sn* become difficult to identify in the records. Instead, *P* and *S* phases show on seismograms. Some workers consider the distance range between 10° and 103° as ideal to record not only the direct *P* and *S* waves, but also the whole family of reflected and converted waves. Travel paths of these waves are dominated by the mantle and corresponding seismograms are relatively simple. *P* is usually stronger on the vertical component, while *S* is more clearly seen on horizontal components. *S* often exhibits wave trains with longer periods when compared with corresponding *P*. Large-amplitude *S* waves are often observed at distances up to about 100°.

A direct *P* reflected from the free surface once or twice is called *PP* or *PPP*, respectively. In the same way we have also *SS* or *SSS*. For distances larger than about 40°, the free-surface reflected phases become very distinct. At distances around 100° and larger, *PP* and *SS* often belong to the largest recorded body waves. Converted waves like *PS* and *SP* appear only at distances larger than 40°. Examples of seismograms with free-surface reflections are displayed in Figure 4.

Core-reflected phases, that is, waves reflected back into the mantle from the outer core-mantle boundary are labeled *PcP* and *ScS*. Including mode converted waves, we have also *PcS* and *ScP*. Because core-reflected phases emerge steeply, *ScP* is usually stronger on vertical component records than *PcS*. Large core-reflected phases are usually recorded at shorter epicentral distances, say at 40° or less. At distances around 39°, *ScP* and *PcS* (surface foci) are often contaminated with the arrival of direct *S* and the phase separation is difficult. When the ray path of *PcP* grazes the outer core boundary, the combination of direct *P* and *PcP* is called *P* diffracted and labeled *Pdif* (older *Pdiff*).

Waves ascending from the focus to the free surface (in continents) or to the sea bottom (under oceans), where they are reflected back into the mantle, are commonly called depth phases. We can list the four possibilities of reflections near the epicenter, which are *pP*, *sP*, *pS*, and *sS*. The first case, for example, denotes the wave that traveled upward from the focus as *P* (short leg) and had been reflected back off the free surface again as *P* (long leg). Depth phases, primarily *pP*, are the most important phases routinely used in focal-depth estimations. The deeper the focus, the later is the *pP* phase in relation to direct *P*. Thus, accurately measured arrival-time differences, *pP-P*, are reliable indicators of the depth of the focus. Most teleseismic shallow events (depth of focus around 33 km) will reveal a *pP* 9–11 s after *P* (see, e.g., Herrin et al., 1968). In case of a deeper focus, it is sometimes possible to recognize several different reflections/conversions from the free surface, for example, *pPP*, *sPP*, *pPS*, *sPS*, *pSP*, etc. Interpretation of depth phases must be done with utmost care since, for example, *pP* from a deep event can easily be erroneously interpreted as *P* when the first arrival (*P*) is weak. Similar difficulty emerges for multiple shocks. Depending on focal orientation and other factors, *sP* may be stronger than *pP* and may be mistaken for it. Depth phases are sometimes stronger than the direct *P*, and may be the first readable phase. We usually require several records made at different epicentral distances to make a reliable identification of the depth phase. We can



Seismogram Interpretation, Figure 4 The Costa Rica earthquake of August 20, 1999, ($M = 6.9$, $h = 40$ km). Exhibited seismograms are long-period, E–W (*upper trace*) and vertical-component (*lower trace*), records made at KONO (Kongsberg), Norway. The epicentral distance is 84°. There are 100 s between two successive time marks at the bottom of the figure. For this distance range, *SKS* arrives ahead of *S*, but the separation is difficult. Both displayed channels are dominated by a fundamental *LR* wave, developed along a mainly oceanic propagation path.

also mention less common phases like *pmP* and *pwP* (first identified by Mendiguren, 1971), corresponding to Moho underside reflection and to water surface reflection, respectively (Schenk et al., 1989). A thin layer of low-velocity sediments at the sea bottom may have small impedance contrast. Hence, short-period records may show large *pwP* when compared with *pP* reflecting at the sea bottom (Engdahl et al., 1998). *pwP* can easily be misinterpreted as *pP* leading to an overestimation of the focal depth.

Bolt (1982) introduced symbols like *P*400*P* or *P*650*P* to indicate seismic waves reflected at the underside of a secondary discontinuity, in the upper mantle, at a depth of 400 and 650 km, respectively. These phases arrive at the recording station ahead of the expected (calculated) arrival time for the main *PP* phase and are interpreted as reflections from upper mantle discontinuities, that is, as *PdP*. When the early arrival time cannot be explained in terms of known discontinuities as *PdP*, we call these onsets early *PP* or precursors to *PP*.

Body waves traversing the core; recording distances 103° and beyond

Due to the diffraction around the core-mantle boundary (see *Earth's Structure, Global*), amplitudes of direct *P* waves decay dramatically at distances larger than 100°. The short-period *P* reappear consistently on records first at distances of about 140°. The distance range $103° < \Delta < 140°$ is called the shadow zone. On seismograms, *Pdif* waves usually show small amplitudes, emergent or gradual onsets and the energy shifts to longer periods. Long-period *Pdif* are sometimes observed out to distances of 160° or more. *S* waves are affected at the core-mantle boundary in a similar way.

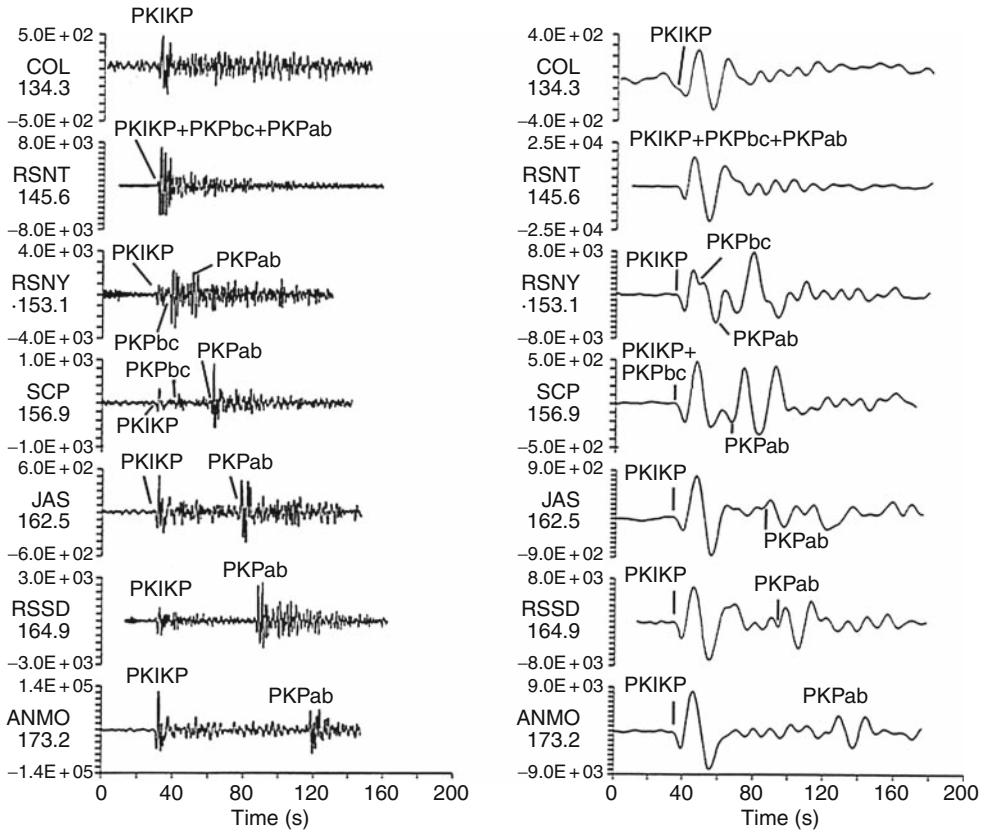
Waves that traveled through the outer core but do not enter the inner core are labeled *PKP* (also *P′*), *PKS*, *SKS* (also *S′*), and *SKP*. The *SKP* phase is stronger on vertical components than *PKS*. These phases have a caustic near 130° and close to this distance are often the only phase recorded on short-period seismograms. At epicentral distances close to 144°, *PKP* shows a distinct concentration of energy, that is, large amplitudes on records. The phenomenon may be viewed in terms of two *PKP* travel-time branches denoted *PKPbc* and *PKPab* for the first and the second arrival, respectively. At 144°, the waves from the two branches coincide, the waves reinforce one another, which results in energy concentration near that distance.

P waves that traverse the inner core are denoted by *I*, giving rise to phases *PKIKP*, *PKIKS*, *SKIKS*, and *SKIKP*, although these are often still simply referred to as *PKP*, *PKS*, etc. Phases, with an *S* leg in the inner core would include the letter *J*, such as *PKJKP*, but these have been difficult to identify on records. Cao et al. (2005) made use of high-quality, broadband seismograms from the Gräfenberg array station, Germany, and identified a phase with arrival time and slowness consistent with predictions of *PKJKP*.

If the studied event is weak, then usually no *Pdif* is observed in the entire distance interval $\Delta > 103°$ and the first arrival seen on the record will be that of *PKP*. At epicentral distances 105–120°, *PKIKP* usually provides the first onset discernible on the seismogram. In the region of the caustic, that is, around 144°, the wave train of recorded core phases becomes particularly complicated. It is first at distances beyond the caustic point where observed onsets may be separated into individual *PKP* branches. The energy distribution changes with the increasing distance. *PKPbc* is the dominant branch just beyond the caustic, up to about 153°. In records of weaker events (Δ = 144–153°), *PKPbc* is often the first visible onset since *PKIKP* (*alt. PKPdf*), theoretically preceding *PKPbc*, is too weak to be observed. As the distance increases, *PKPbc* becomes weaker and vanishes from records at distances of about 160° and larger. For distances beyond, say, 157°, *PKPab* usually dominates the seismogram. *PKIKP* in the distance range from about 125° to the caustic is often preceded by early arrivals or precursors, which can arrive many seconds ahead of the main phase. These are often explained by scattering phenomena at or near the core-mantle boundary. Subscripts *ab*, *bc*, and *df* are used in agreement with travel-time charts of Jeffreys and Bullen (1940). Recorded core phases made at a suite of seismographic stations at a distance range from 134° to 173° are shown in Figure 5.

We may form new symbols for the whole family of waves propagating through the outer core. For example, *PKKP* is a *P* wave reflected from the inside of the core-mantle boundary, often very pronounced on records made at distances between 60° and 80°. However, the striking onset may easily be misinterpreted as a first *P* arrival of another event. *P* waves trapped inside the Earth's liquid core and with multiple *K* legs are called *PNKP*, where $N-1$ is the number of reflections. Multiple reflections within Earth's outer core were first observed by Engdahl (1968). He used records from two deep earthquakes and an event in Novaya Zemlya ($h = 0$) and identified recorded seismic waves reflected as many as four times (i.e., *P5KP*) within the Earth's outer core. Cases like *P4KP* and *P7KP* have been reported by Bolt (1982). Waves that traverse the Earth's interior and are reflected at the outside or inside of the inner core are labeled *PKiKP* or *PKIIKP*, respectively.

The best chance to observe *PKPPKP*, or for short *P′P′*, is around distances $2 \times 144° = 288°$, or 72° if we take the shortest distance from source to station. *P′P′* is often well recorded, arriving about 30 min after *P*, when most of the coda amplitudes of preceding phases have already become faint. *P′P′* may in some cases be wrongly interpreted as a new *P* or *PKP*. 72° is also equivalent to $3 \times 144° = 432°$, so the phase *P′P′P′* is also strong at this distance, and may be observed, for strong earthquakes, about another 20 min after *P′P′*. In 1969, first observations of precursors to *P′P′*, sometimes called early *P′P′*, were made (Engdahl and Flinn, 1969). These were interpreted as *P′dP′* analogous to *PdP*, that is, as reflections of *PKP* at



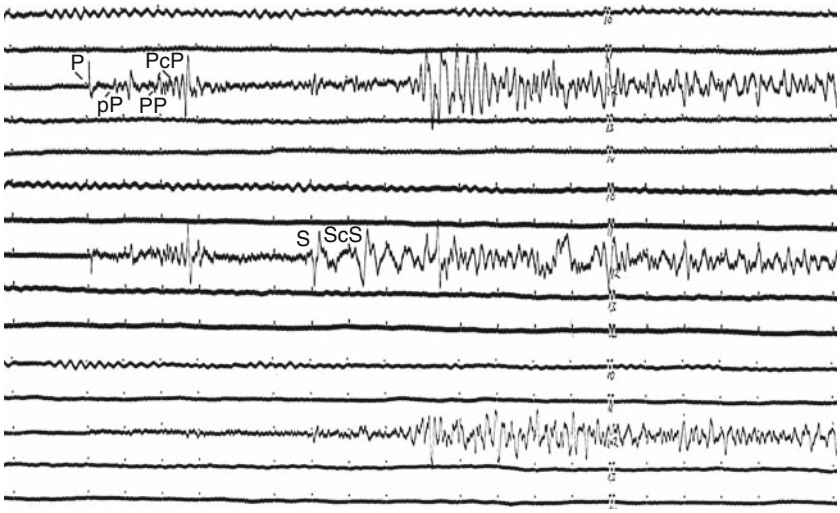
Seismogram Interpretation, Figure 5 Seismograms from a Mid-Indian-Rise earthquake of May 16, 1985 (M = 6.0, h = 10 km) made at a suite of seismographic stations that sample the epicentral distance between 134.3° and 173.2°. Short-period (*left*) and long-period (*right*), vertical-component records are exhibited. Station codes, epicentral distances, and amplitude scaling are given to the left of each trace (G. Choy, personal communication; reproduced from Kulhanek, 1990, with permission from Elsevier, Science).

secondary velocity discontinuities within Earth's upper mantle. For example, $P'650P'$ passes through the core twice and on seismograms precedes $P'P'$ by about 2 min.

Similar to *PKP* and *PNKP*, there are *SKS* and *SNKS* waves, respectively. First *SKS* waves are observed at distances between 60° and 70° and the range of observations extends to distances of 180° or so. *SKS* exhibits a caustic point at a distance of about 80° so that the best region to study *SKS* waves is that between 70° and 90°. The phase identification has to be made with utmost care since *SKS* recorded at this distance range are often contaminated with direct *S*. To mistake *S* for *SKS* and vice versa will adversely affect the epicentral location. At about 82° (depending upon details in the structural model), *SKS* begins to arrive ahead of *S* (Figure 4). For distances shorter than about 95°, *SKS* is usually smaller than *S*, however, at distances beyond 95°, *SKS* amplitudes are quite large. *S* and *SKS* are best recorded on long-period horizontal-component seismograms. The period of *SKS* phases may reach several tens of seconds. Occasionally, these body waves are also seen on short-period records, although the onset time of the later of the two phases is usually very emergent due to the contamination by the coda of the earlier phase.

Body waves from intermediate-focus and deep-focus earthquakes

Later studies confirmed conclusions from Wadati's pioneering work (Wadati, 1927) that intermediate-focus and deep-focus shocks produce simpler seismograms with exceptionally well recorded impulsive body waves while surface-wave amplitudes decrease as the quake becomes deeper. Strong depth phases, such as *pP* and *sS*, are also frequently very distinct on records from deep events. However, the duplication of principle phases by surface reflection (e.g., confusion between *pP* and *PP*) often complicates the seismogram interpretation. Weak precursors, *pmP*, (Moho underside reflections) to surface reflections, *pP*, are discussed in Schenk et al. (1989). They are best observed in long-period records and in continents often produce clear arrivals. In a more retrospective-type interpretation, the absence of aftershocks (see *Earthquake, Aftershocks*) will support the classification of the shock as



Seismogram Interpretation, Figure 6 Seismograms from an intermediate-focus earthquake (M = 6, h = 199 km) in the Hindu-Kush region on September 4, 1993, made at UPP, Sweden, at an epicentral distance of 41°. The traces show long-period channels from top to bottom: vertical, E–W and N–S components.

a deep-focus or intermediate-focus event. Figure 6 shows seismograms from an intermediate-focus earthquake made at an epicentral distance of 41°. Note the impulsive character of the *P*, *pP*, *PcP*, *S*, and *ScS* phases; virtually no surface waves have been recorded from this event.

Surface waves

With exception of very short epicentral distances, long-period surface waves, *LQ*, and *LR* dominate the seismograms of shallow events. Deep events of the same magnitude will generate abnormally small (insignificant) surface waves. This feature provides the interpreter with a viable tool to discriminate, at first glance, between shallow and deep-focus earthquakes. (Note a paradox: underground nuclear explosions, being very shallow events, generate only small, sometimes invisible, surface waves). The analyst has several clues to identify surface waves on the record and to distinguish *LQ* from *LR*. While Love waves are best displayed on horizontal-component seismograms, Rayleigh waves are best seen on vertical-component records. Both *LQ* and *LR* propagate slower than *P* or *S*, but since *LQ* propagate faster than *LR*, they are recorded ahead of *LR*. Due to dispersion, long-period surface waves advance on records toward the beginning of the wave train as it travels through the medium. An experienced analyst will distinguish between recorded surface waves that have traveled along pure oceanic or continental path. Dispersion of oceanic routes give rise to long wave trains with rather slow and sometimes hardly visible period change over relatively long (10–20 min) record segments (see Figure 4). In contrast, continental paths generate surface waves with characteristic fast period decrease with time, which is often easily recognized on records (see Figure 1). Surface waves can also travel by different modes (overtones), which are often seen on records as high-frequency components superimposed on the surface-wave train. We talk then about fundamental-mode and higher-mode surface waves (see *Surface Waves*). Higher modes are most frequently observed for waves traversing purely continental paths. Nevertheless, higher modes have in some cases also been associated with oceanic paths. Surface-wave higher modes disappear when the waves cross the transition between continental and oceanic structures. Higher modes propagate faster than the fundamental mode and are, therefore, recorded ahead of *LQ* and *LR*.

Dispersion curves (see *Surface Waves*) show a rather complicated pattern with several local minima and maxima. Surface waves traveling with these minima or maxima group velocities are called Airy phases. On seismograms, an Airy phase is characterized by a constant-frequency compact wave train, often with a remarkable amplitude buildup of dispersed surface waves traveling by fundamental-mode propagation. In the period range from approximately 50–200 s, the group velocity of *LR* (both the oceanic and continental paths) monotonically decreases with increasing period. Physically this means that in this period range, long-period *RL* waves follow the law of inverse dispersion. Observations of this phenomenon are rather scarce. An example is exhibited in Kulhanek (1990).

Extremely long-period surface waves called mantle waves, with periods from somewhat less than 1 min to several minutes, have been observed from large distant shocks. They can be of either Love- or Rayleigh-wave type. The former propagates with nearly constant speed of 4.4 km/s and shows an impulsive shape on the seismogram. The latter travels with velocity between 3.6 and 4.1 km/s. Wavelengths of mantle waves vary from several

hundreds to more than 1,000 km, so that a large part of Earth's mantle is affected by these waves. An interesting feature of mantle waves is their repeated appearance on records, which is due to their multiple travel around the Earth. The *LQ* mantle wave was given the label *G* (after B. Gutenberg) and the *LR* type mantle wave, the label *R*. The older nomenclature sometimes uses *W* instead for *R* (from German *Wiederkehrewellen*, meaning repeated waves). *G* waves that propagate the direct and anticenter routes are labeled *G1* and *G2*, respectively. Waves that have in addition traveled once around the Earth are denoted *G3* and *G4*, and so on. Similarly, we have *R1*, *R2*, *R3*, *R4*, etc. As an exceptional case we can mention records of the 1960 Chile earthquake, $M = 9.5$. Seismograms made at Uppsala, Sweden, reveal mantle waves *G20* and *R20* that have traveled a total distance equal to that from the Earth to the Moon (Båth, 1979).

Volcanic earthquakes and unusual seismic sources

During eruptive episodes, volcanoes can produce up to thousands of small earthquakes per day, recorded only at short distances, say, less than 50 km, from the volcanoes (see *Earthquakes, Volcanogenic*). Different categorization of volcanic earthquakes may be found in the literature. For example, Minakami (1961) classifies volcanic earthquakes into three groups, in accordance with different generation mechanisms and signal characteristics. (1) A-type earthquakes with focal depth between 1 and 10 km near the volcanic magma system. Corresponding records show clear high-frequency *P* and *S* phases. (2) B-type earthquakes with foci at depth of 1 km or less. Records reveal low-frequency coda, emergent onsets, and usually no distinct *S*. (3) Explosion-type earthquakes taking place at the very surface of the Earth. Close to active volcanoes, we also frequently detect so-called volcanic tremors, which are due to long-duration, more or less continuous, volcanic vibration.

Implosion earthquakes, impact earthquakes (e.g., the Tunguska, Siberia, impact in 1908), frost actions, low-magnitude icequakes generated by temperature changes in glaciers, earthquakes related to large-scale landslides, etc. are some of the types of events that, together with tectonic earthquakes, volcanic earthquakes, and oceanic microseisms, belong to the category of natural seismic sources. There is also a variety of man-made seismic sources such as industrial and military explosions, cultural noise (traffic, industry work), mining activity, high dams (seismicity, triggered/induced), fluid injections, etc. The source identification, for earthquakes other than tectonic or volcanic, is usually a task in itself. A classical example of source identification is the well-known problem of discriminating underground nuclear explosions from earthquakes (see *Seismic Monitoring of Nuclear Explosions*).

Conclusions

Seismogram interpretation, described in this article, is essentially devoted to the art of identification of various seismic "arrivals" or wave types visible on seismograms. It is usually followed by seismogram analysis, which may include determination of basic source parameters (origin time, hypocenter location, size) but may also cover more advanced studies, such as wave-form modeling, determination of velocity distribution, etc. Obviously, a large part of seismogram analysis is a domain of research. Nevertheless, phase identification is a doorway and without correct seismogram interpretation hardly any analysis would be possible. Repeated observations of peculiar phases on seismic records often led to new discoveries of details in Earth's structure and/or dynamics. Sometimes, theories were developed in advance, to be later confirmed by observations (e.g., free oscillations of the Earth). Seismogram interpretation is a fundamental, and in our view, also fascinating and rewarding part of modern seismology.

Bibliography

Adams, R. D., 1979. T-phase recordings at Rarotonga from underground nuclear explosions. *Geophysical Journal of the Royal Astronomical Society*, **58**, 361–369.

Bolt, B. A., 1982. *Inside the Earth*. Berkeley: Freeman.

Bormann, P. (ed.), 2002. *New Manual of Seismological Observatory Practice (NMSOP)*. Potsdam: Deutsches GeoForschungsZentrum.

Båth, M., 1979. *Introduction to Seismology*. Basel/Boston/Stuttgart: Birkhäuser.

Båth, M., 1983. Earthquake data analysis: An example from Sweden. *Earth-Science Review*, **19**, 1891–303.

Cao, A., Romanowicz, B., and Takeuchi, N., 2005. An observation of *PKJKP*: Inferences on inner core properties. *Science*, **308**, 1453–1455.

Engdahl, E. R., 1968. Seismic waves within Earth's outer core: Multiple reflections. *Science*, **161**, 263–264.

Engdahl, E. R., and Flinn, E. A., 1969. Seismic waves reflected from discontinuitie within Earth's upper mantle. *Science*, **163**, 177–179.

Engdahl, E. R., van der Hilst, R. D., and Buland, R., 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America*, **88**, 722–743.

Herrin, E., Arnold, E. P., Bolt, B. A., Clawson, G. E., Engdal, E. R., Freedman, H. W., Gordon, D. W., Hales, A. L., Lobdell, J. L., Nuttli, O., Romney, C., Taggart, J., and Tucker, W., 1968. Seismological Tables for P phases. *Bulletin of the Seismological Society of America*, **58**, 1193–1352.

Jeffreys, H., and Bullen, K. E., 1940. *Seismological Tables*. London: British Association for the Advancement of Science.

Kennett, B. L. N., 1991. *IASPEI 1991 Seismological Tables*. Canberra: Research School of Earth Sciences, Australian National University.

Kennett, B. L. N., and Engdahl, E. R., 1991. Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, **105**, 429–465.

Kennett, B. L. N., Engdahl, E. R., and Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, **122**, 108–124.

Kulhanek, O., 1990. *Anatomy of Seismograms*. Amsterdam: Elsevier.

Mendiguren, J. A., 1971. Focal mechanism of a shock in the middle of the Nazca plate. *Journal of Geophysical Research*, **76**, 3861–3879.

Minakami, T., 1961. Study of eruptions and earthquakes originating from volcanoes, I. *International Geology Review*, **3**, 712–719.
Morelli, A., and Dziewonski, A. M., 1993. Body wave traveltimes and a spherically symmetric P- and S-wave velocity model. *Geophysical Journal International*, **112**(178), 194.
Neumann, F., 1951. Principles underlying the interpretation of seismograms. U.S. Department of Commerce, Coast and Geodetic Survey. Special publication number, 254.
Payo, G., 1986. *Introduccion al Analisis de Sismogramas*. Madrid: Instituto Geográfico Nacional.
Schenk, T., Muller, G., and Brustle, W., 1989. Long-period precursors to *pP* from deep-focus earthquakes: the Moho underside reflection *pMP*. *Geophysical Journal International*, **98**, 317–327.
Simon, R. B., 1968. *Earthquake Interpretations*. Golden: Colorado School of Mines.
Storchak, D. A., Borman, P., and Schweitzer, J., 2003. The IASPEI Standard Seismic Phase List. *Seismological Research Letters*, **74**, 761–772.
Wadati, K., 1927. Existence and study of deep earthquakes (in Japanese). *Journal of the Meteorological Society of Japan*, Ser 2, 5, 119–145.
Willmore, P. L., 1979. *Manual of seismological observatory practice*. Report SE-20, World Data Center A for Solid Earth Geophysics, NOAA, Boulder.

Cross-references

Body Waves
Core-Mantle Coupling
Earth's Structure, Upper Mantle
Earthquake, Aftershocks
Earthquake, Focal Mechanism
Earthquakes, Volcanogenic
Seismic Monitoring of Nuclear Explosions
Seismic Phase Names: IASPEI Standard
Seismic Velocity-Density Relationships
Surface Waves
Traveltime Tomography Using Controlled-Source Seismic Data

SEISMOLOGICAL NETWORKS

Eric Robert Engdahl[1], István Bondár[2]
[1]Center for Imaging the Earth's Interior, Department of Physics, University of Colorado at Boulder, Boulder, CO, USA
[2]International Seismological Centre, Pipers Lane, Thatcham, Berkshire, UK

Definition

Seismological Networks. Networks of seismographic stations for the recording of ground motions produced by seismic waves propagating from natural and anthropogenic sources.

Introduction

Seismology has its roots in observations of earthquakes and earthquake-generated ground motion. Seismological networks as envisioned by the pioneers in seismology are now a reality, a reality that is frequently upgraded and expanded as technology improves. A global system of broadband seismographs with high dynamic range is now in place. This is supplemented by many national, regional, and local networks capable of high-resolution monitoring of the frequent smaller earthquakes in seismically active places. Parametric data derived from these networks make a major contribution to national and international information services. Because of page limitations, we can only hope to provide the reader with a glimpse of all there is to know about seismographic networks.

Global seismic networks

Seismological recordings have been made on Earth for hundreds of years in some form or another. However, global monitoring of earthquakes only began in the 1890s when John Milne created 40 seismic observatories to measure the waves from these events (see Figure 1). Shortly after the International Geophysical Year (1957–1958), a concerted effort was made to establish and maintain a more modern standardized seismic network on the global scale.

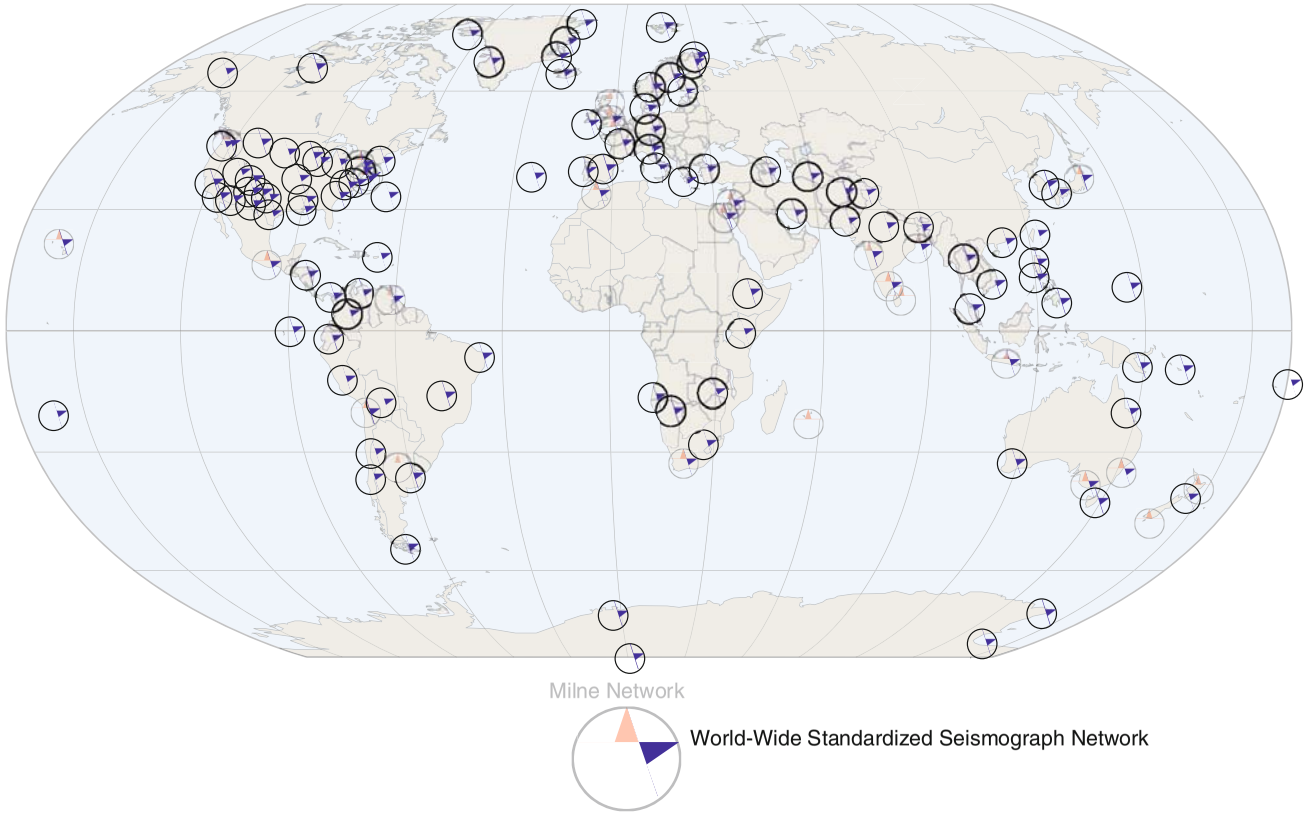
Worldwide Standardized Seismograph Network

In the early 1960s, the World-Wide Standardized Seismograph Network (WWSSN) was established. Between 1961 and 1966, 120 WWSSN stations with identical seismic instruments were deployed in more than 60 countries and islands throughout the world. The WWSSN program included the collection, review, and copying of the recorded seismograms, and the distribution of copies to researchers throughout the world. Considered one of the most important advances ever in observational seismology, the WWSSN produced the data needed to support unprecedented progress in earthquake, explosion, and tectonic research. In the years that followed the deployment of the WWSSN, many of the stations were modernized and expanded into regions not initially covered. Eighty of the original WWSSN stations are still in operation and still supported.

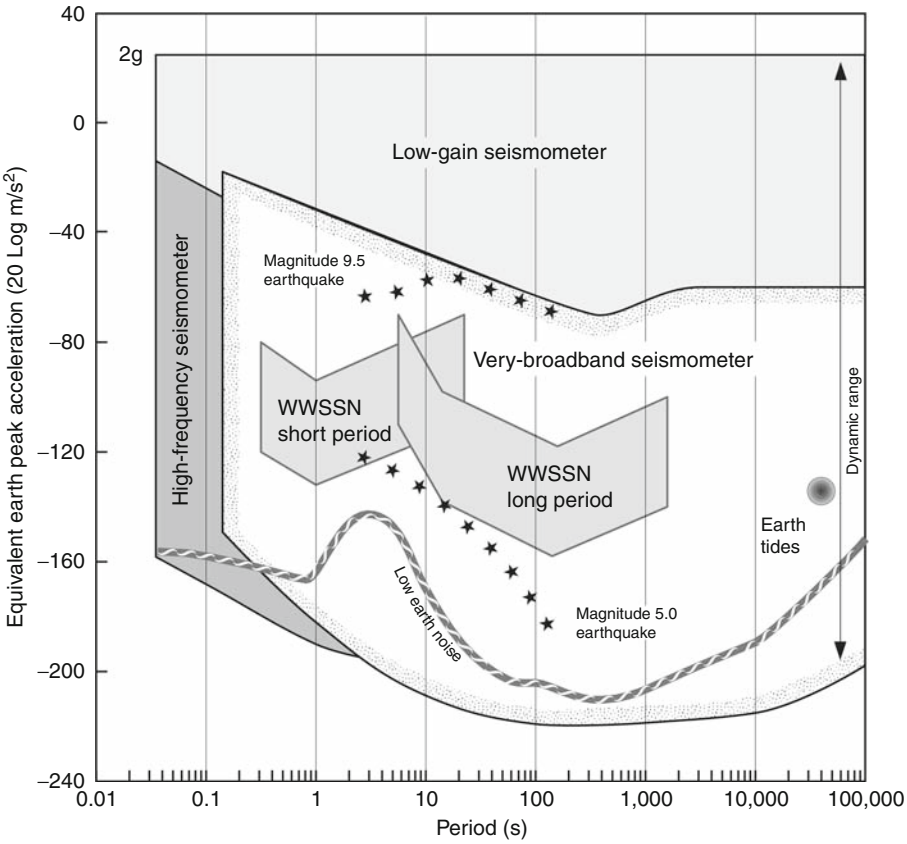
International Federation of Digital Seismograph Networks

In the 1980s, the international seismological community recognized the new opportunities within its field for improved understanding of the internal structure and dynamical properties of the Earth provided by recent developments in seismograph network technology. It also recognized that rapid access to seismic data networks of modern broadband digital instruments wherever they might be was now possible. The developments included greatly improved broadband seismographic systems that capture the entire seismic wave field with high fidelity (see Figure 2), efficient and economical data communications and storage, and widely available, powerful computing facilities.

In view of the above and to take advantage of existing developing global and regional networks, the International Federation of Digital Seismograph Networks (FDSN, http://www.fdsn.org/) was formed to provide a forum for: developing common minimum standards in seismographs (e.g., bandwidth) and recording characteristics (e.g., resolution and dynamic range); developing



Seismological Networks, Figure 1 Locations of stations in the Milne network and the World-Wide Standardized Seismograph Network. (Courtesy of KentAnderson, Incorporated Research Institutions for Seismology.)



Seismological Networks, Figure 2 Bandwidth and dynamic range of WWSSN and modern seismographic systems. (Courtesy of Rhett Butler, Incorporated Research Institutions for Seismology.)

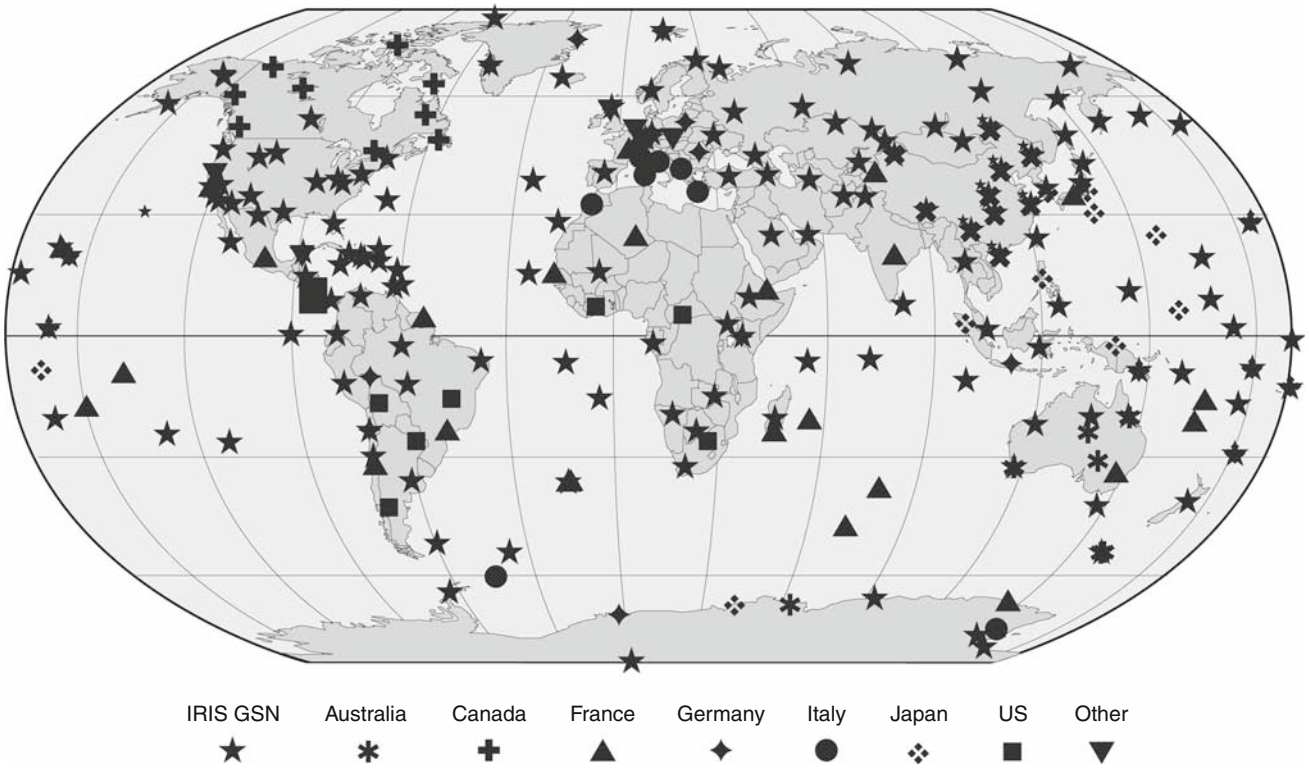
standards for quality control and procedures for archiving and exchange of data among component networks; and coordinating the siting of additional stations in locations that will provide optimum global coverage.

The FDSN was formed as a global organization, open to all national and international programs committed to the deployment of broadband seismographs and willing to contribute to the establishment of an optimum global system with timely data exchange. Its membership is comprised of groups responsible for the installation and maintenance of seismographs either within their geographic borders or globally. Membership in the FDSN is open to all organizations that operate more than one broadband station. However, among the most important contributors is the Global Seismographic Network (GSN) of the Incorporated Research Institutions for Seismology (IRIS), and the French GEOSCOPE and German GEOFON global networks. It is important to note that many stations in the FDSN were formerly part of the WWSSN program. Members agree to coordinate station siting and provide free and open access to their data. This cooperation helps scientists all over the world to further the advancement of earth science and particularly the study of global seismic activity.

The FDSN goals related to station siting and instrumentation are to provide stations with good geographic distribution, recording data with 24 bits of resolution in continuous time series with at least a 20 sample per second sampling rate. The FDSN was also instrumental in development of a universal standard (SEED) for distribution of broadband waveform data and related parametric information. The FDSN system of global network observatories includes contributions from many international partners (Figure 3), forming a large backbone of permanent seismological observatories. Developments in communications and other technological advances have expanded the role of the FDSN in rapid earthquake analysis, tsunami warning, and nuclear test monitoring. With such long-term observations, scientists are now getting a glimpse of Earth structure changes on human time scales, such as the rotation of the inner core. Continued observations for the next 50 years will enhance our image of the Earth and its processes.

International Monitoring System

Under the Comprehensive Test-Ban Treaty (CTBT, http://www.ctbto.org/), it is a requirement that there will be an International Monitoring System (IMS) to detect any clandestine nuclear weapon detonation in any environment – underground, under-water, or above ground. The aim of



Seismological Networks, Figure 3 Locations of backbone stations in the International Federation of Digital Seismograph Networks with major partners indicated by *symbols*. (Courtesy of Rhett Butler, Incorporated Research Institutions for Seismology.)

the IMS is detection, identification, and location of any such test, and the technologies involved will be seismological, hydroacoustic, infrasound, and atmospheric radionuclide monitoring. In support of these technologies, there will be appropriate means of global communication and an International Data Centre (IDC) to which monitoring data will be transmitted.

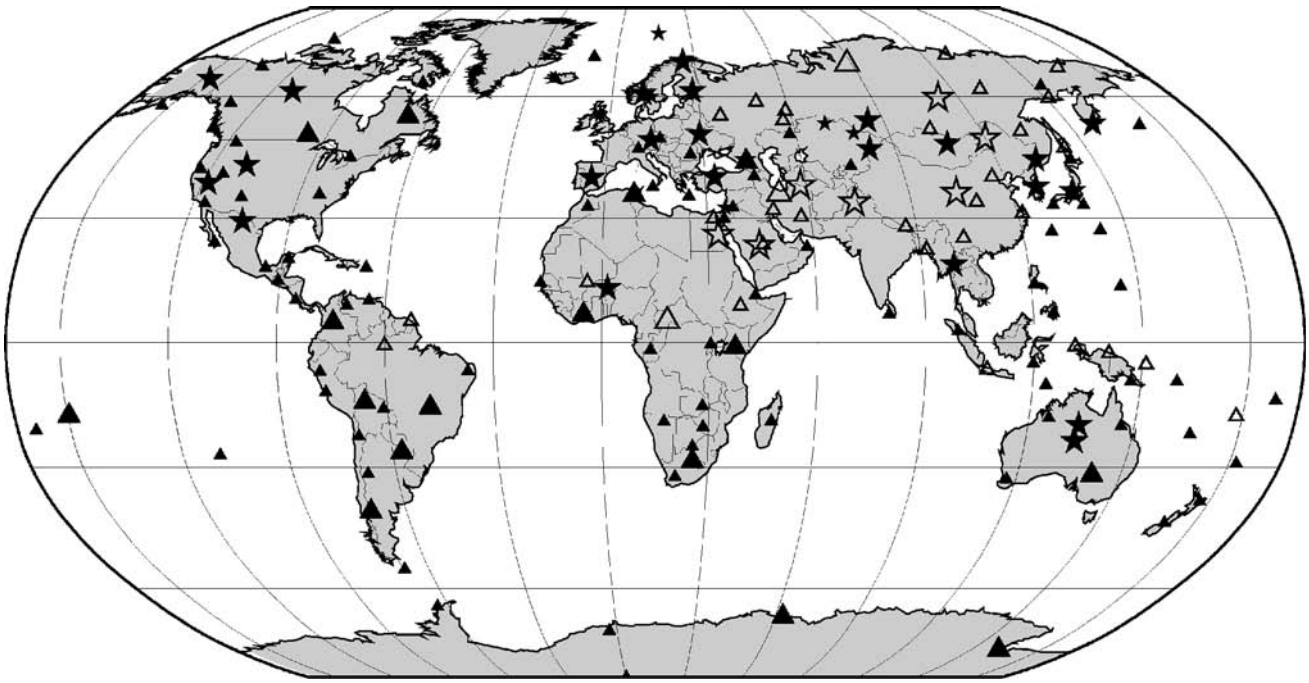
Seismological monitoring is considered the cornerstone of the IMS because of the likelihood that tests would be conducted underground and out of sight. The seismological technology required for the detection of earthquakes is well known, but the problem for the IMS is to distinguish between signals from weapons detonations, earthquakes, and non-nuclear explosions in quarrying or mining. There are so many earthquakes (>200,000 per year) and mining events worldwide, of similar magnitude to a small nuclear explosion, that data analysis becomes a key problem. The seismological verification program therefore includes intensive study of seismicity and mining operations in order to develop regional backgrounds.

A global network of 50 "primary" stations in 34 countries is planned, with two basic equipment packages: "three-component" (3-C) broadband seismometers, which act as single detection systems monitoring ground movement in 3 directions – vertically, and horizontally east-west and north-south; or "arrays" which include 3-C systems plus a cluster of narrowly spaced vertical-component short-period sensors distributed to gain optimum signal-to-noise ratios by waveform correlation. Arrays may contain many detectors located over large areas (up to 200 km^2). The advantage of an array station is that it allows approximate event location on a stand-alone basis. Important historical developments such as the LASA, NORSAR, Graefenberg arrays, and more recently the wide but densely distributed USArray, have marked important steps in seismological network development that have facilitated new research fields and services in seismology. As Figure 4 indicates, approximately half of the proposed stations will be arrays.

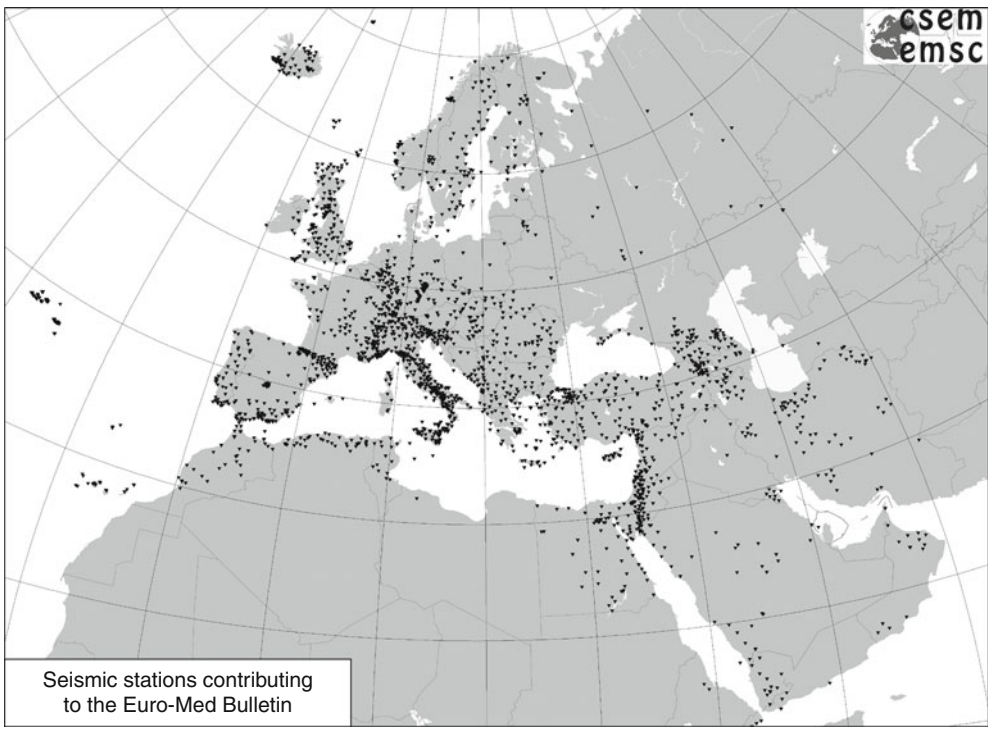
In addition to the primary network, there is to be an "auxiliary" network comprising 120 stations distributed among 61 countries (Figure 4). Stations in this network will mainly consist of existing 3-C stations, which already form part of the host countries' seismological monitoring operations. The purpose of the auxiliary network is simply to provide additional data to support that from the primary network in order to facilitate signal discrimination and hence event detection and location.

Regional seismic networks

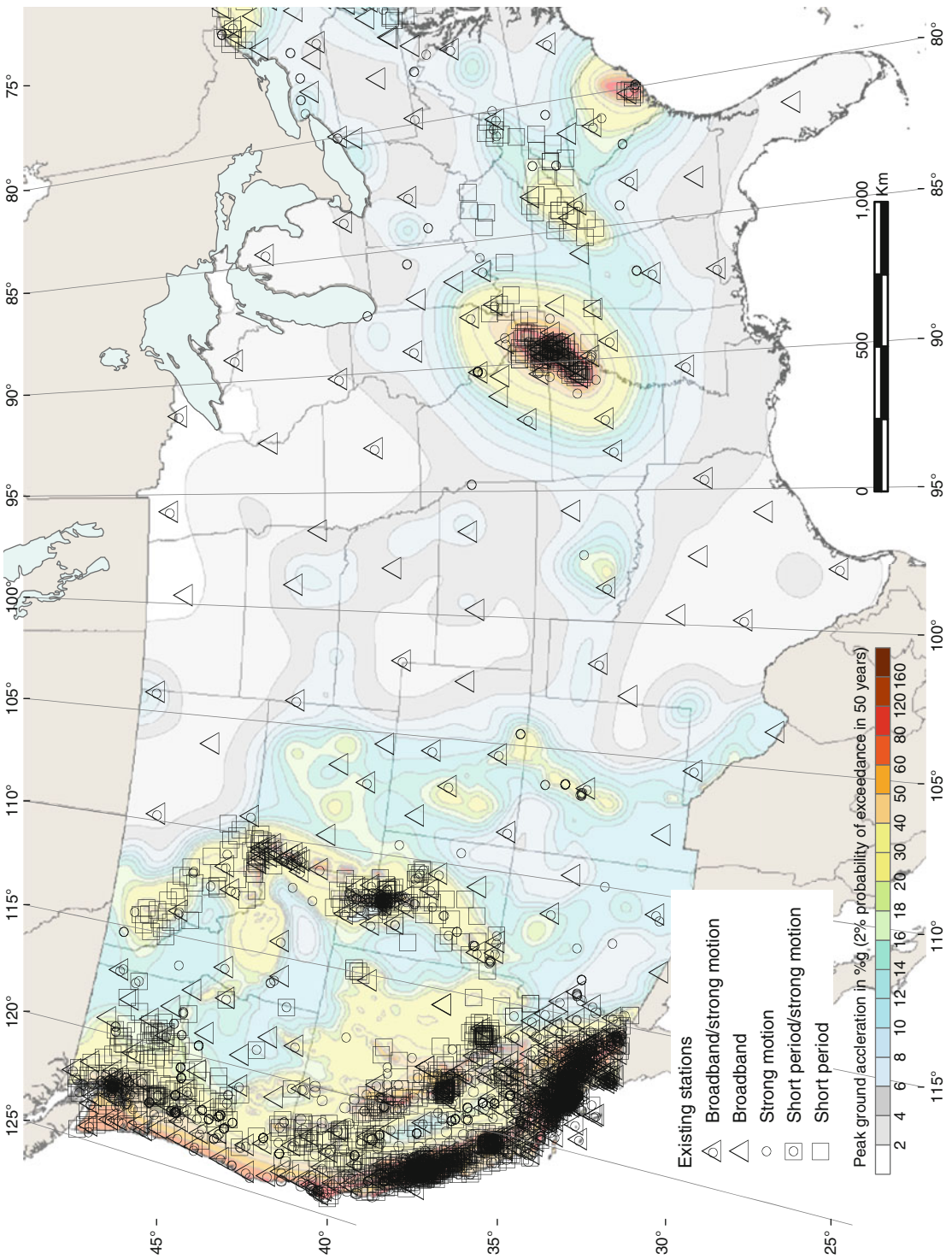
National seismographic networks operate within the political boundaries of a country, and their primary mission is to issue rapid alerts to government agencies and the general public for potentially damaging earthquakes. National



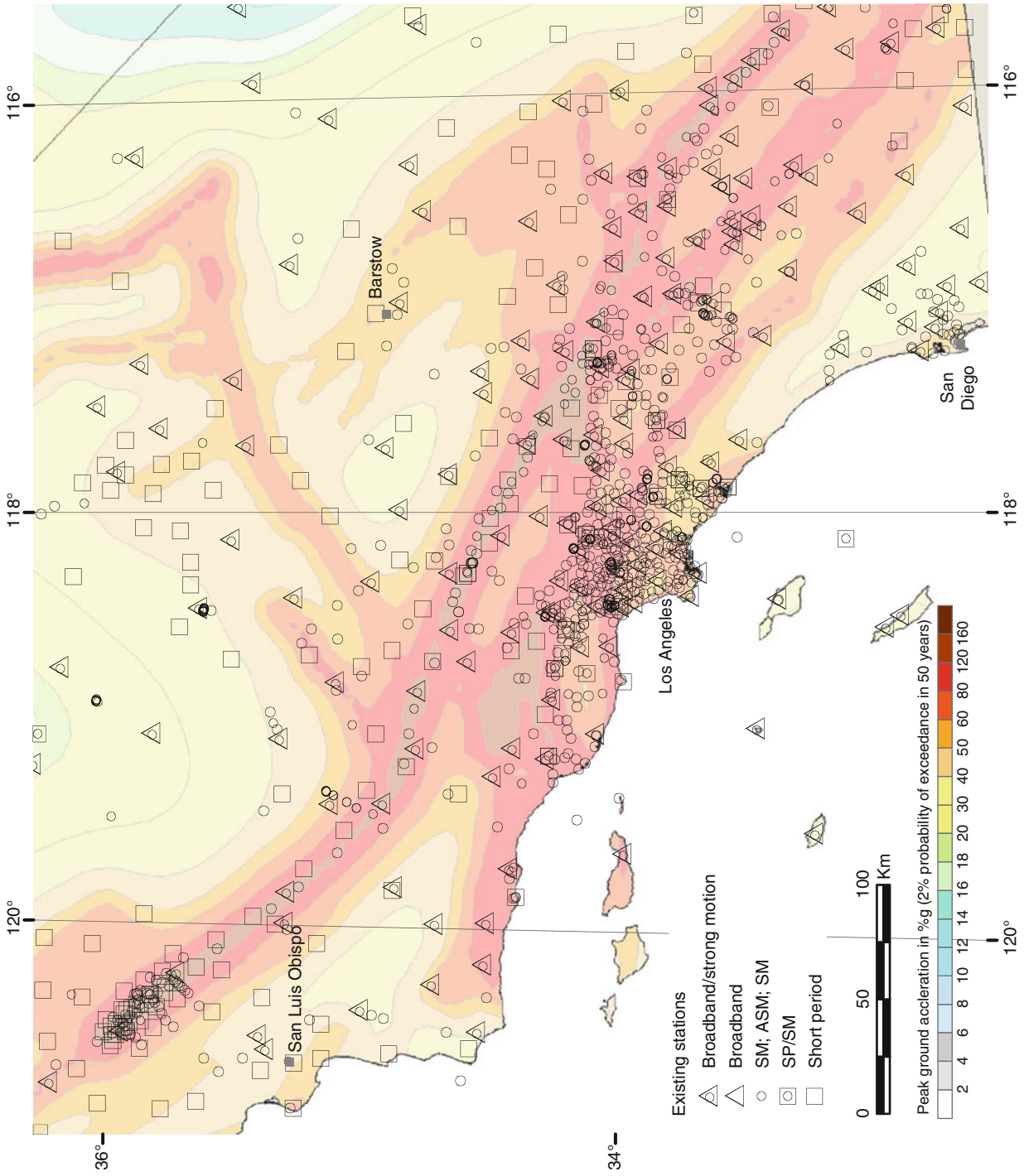
Seismological Networks, Figure 4 Locations of stations in the CTBTO IMS network (Courtesy of Ronan LeBras, Comprehensive Test-Ban-Treaty Organization). Symbols defined as follows: filled symbols – operational; open symbols – not yet operational; *stars* – array stations: *triangles* – three-component stations; *large symbols* – primary stations; and *small symbols* – auxiliary stations.



Seismological Networks, Figure 5 Locations of the contributing stations to the Euro-Med Bulletin 1998–2007 (Courtesy of Stephanie Godey, European-Mediterranean Seismological Centre).



Seismological Networks, Figure 6 Permanent seismic stations in the 48 conterminous states of the United States. (Courtesy of Susan Rhea, U.S. Geological Survey, National Earthquake Information Center.)



Seismological Networks, Figure 7 The Southern California Seismic Network: an example for a very dense regional network. (Courtesy of Susan Rhea, U.S. Geological Survey, National Earthquake Information Center.)

networks typically exchange data with neighboring countries in order to improve the accuracy of the determination of earthquake parameters in border regions.

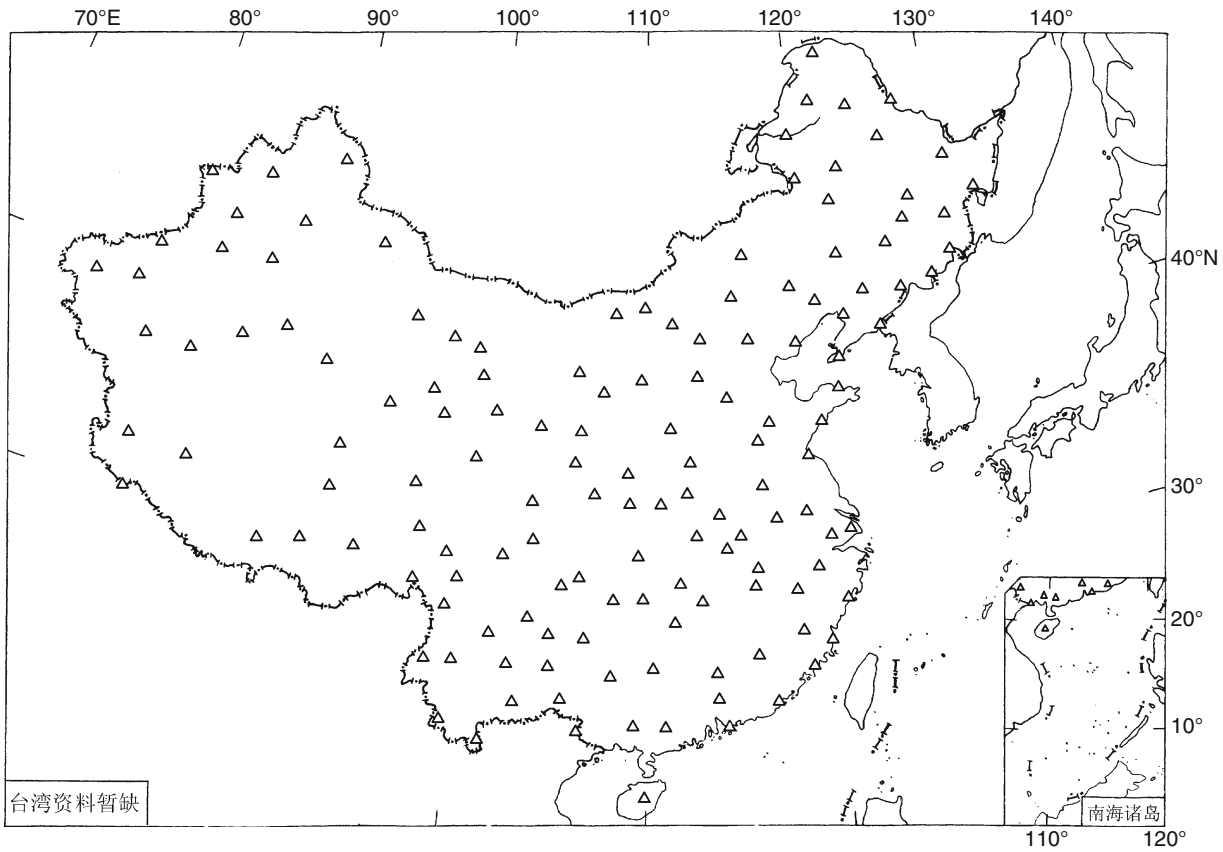
Regional networks play an important role in monitoring the seismicity of the Earth within a well-defined region. Regional networks may be part of a national network (e.g., USA, China, Japan, and Russia) or may be constituted from multinational networks (e.g., Euro-Mediterranean region). The magnitude detection threshold of regional networks is typically lower than that of the global networks, which allows for producing bulletins for a specific region that are more complete than it could be achieved at a global scale.

Local networks focus on a much smaller area than regional networks. Examples of local networks include temporary deployments to capture aftershock sequences of a large earthquake, networks monitoring volcano activities, and dense local networks installed in urban areas with elevated seismic risk.

Euro-Med region

The Euro-Med region encompasses Europe, North Africa, and the Middle East. The European-Mediterranean Seismological Centre (CSEM/EMSC, http://www.emsc-csem.org) produces earthquake bulletins in the region since 1998, using parametric data reported by national and local networks in the region. Its mission is to produce a rapid comprehensive seismological bulletin and issue alerts for potentially damaging earthquakes in the region. The EMSC and the National Earthquake Information Center (NEIC) routinely exchange data, and NEIC bulletin data for the European-Mediterranean region are incorporated in the EMSC bulletin. Earthquake parameters (location, magnitude, phase picks, moment tensors, etc.) provided by the EMSC are incorporated in the global bulletin prepared by the International Seismological Centre (ISC).

Figure 5 shows the seismic station network that currently contributes to the EMSC bulletin (Godey et al., 2006; Godey et al., 2009). All stations are registered at the International Registry of Seismograph Stations jointly maintained by the ISC and the NEIC. The recent years saw a dramatic increase in the number of stations in the region, especially in North Africa and the Middle East. These networks are vital for improving the azimuthal station coverage for the events, and thus improving their location accuracy. Waveform data from most stations contributing to the EMSC can be obtained from Observatories and Research Facilities for European Seismology (ORFEUS)



Seismological Networks, Figure 8 Locations of stations belonging to the China Digital Seismograph Network. (Courtesy of Liu Ruifeng, China Earthquake Network Center.)

the coordinating organization for seismic networks in Europe (http://www.orfeus-eu.org).

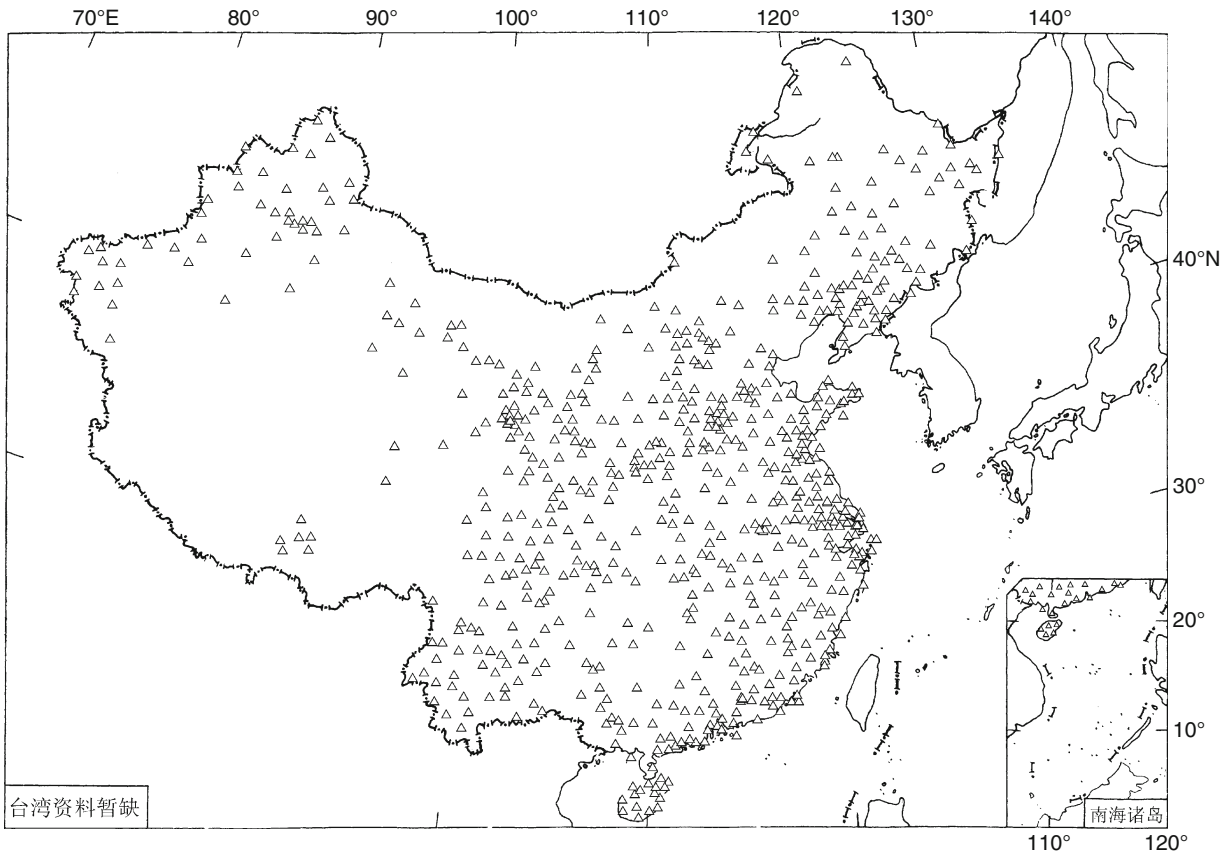
United States

In the United States, the U.S. Geological Survey's National Earthquake Information Center (NEIC, http://earthquake.usgs.gov/regional/neic) is the national data center and archive for earthquake information. Its mission is to determine as rapidly and accurately as possible the location and size of all significant earthquakes that occur worldwide. The NEIC operates 24-h-a-day and issues rapid alerts for earthquakes larger than magnitude 3 in the conterminous states of the United States and for those larger than magnitude 5 globally. The NEIC publishes global earthquake bulletins on a daily, weekly, and monthly basis.

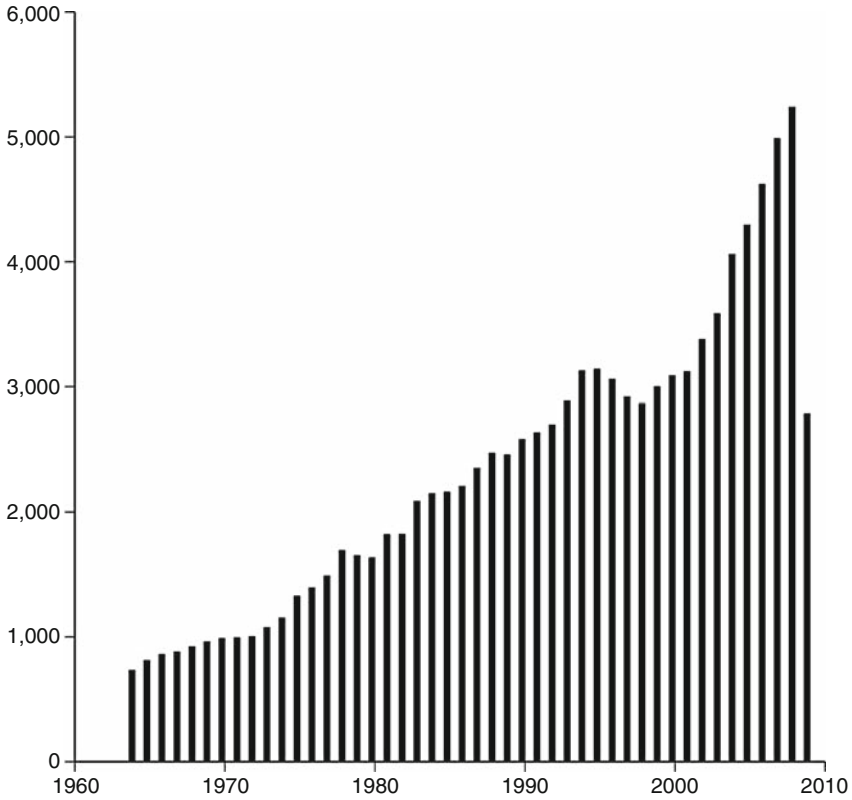
The NEIC serves as the National Operations Center of the Advanced National Seismic System (ANSS), which includes the ANSS backbone network and regional seismic networks across the United States. Figure 6 shows the permanent digital seismic and strong motion stations in the 48 conterminous states of the United States. The strong motion stations (circles) represent an integral part of the ANSS network as first arriving phases are picked on strong motion records and used in the earthquake location procedures. Figure 7 shows the regional Southern California Seismic Network. As in Figure 6, the map is colored by the peak ground acceleration, a measure of seismic hazard. The network is much denser in urban areas and regions with high seismic hazard.

China

The China Digital Seismograph Network (CDSN) has been established in 1986, with nine digital seismic stations. During the past 10 years, the Chinese seismic network, both at the national and regional levels, went through unprecedented improvements. The China Earthquake Administration (CEA) has completed the analog-to-digital transition of the existing analog stations and deployed a large number of new digital stations (Liu et al., 2008). Figure 8 shows the currently operating national network of 145 broadband seismographic stations. Some of these stations also belong to the FDSN. Dense regional networks further support the national network. Each of the 31 provinces, autonomous regions, and municipalities operates regional digital networks. Figure 9 shows the distribution of the 792 regional seismic stations.



Seismological Networks, Figure 9 Digital seismographic stations operated by 31 regional networks in China. (Courtesy of Liu Ruifeng, China Earthquake Network Center.)



Seismological Networks, Figure 10 Number of stations reporting to the ISC between 1964 and 2009. (Courtesy of James Harris, International Seismological Centre.)

The National Earthquake Network Center (NENC) archives the continuous waveform data from the national and regional networks and performs real-time data processing. The NENC issues rapid earthquake alerts for Ms >4.5 events inside China within 10 min and focal mechanism solutions in no more than 30 min after the earthquake occurred.

International Registry of Seismograph Stations

Since the humble beginnings of instrumental global seismology in the late nineteenth century, the number of seismographic stations has steadily increased every year. Figure 10 shows the number of stations reporting to the ISC in each year since its operations began. In order to be able to uniquely identify seismographic stations, the ISC and the NEIC, in its capacity as World Data Center A for Seismology, jointly maintain the International Registry of Seismograph Stations (IRSS). The IRSS represents a global catalog of seismic stations (currently with some 15,000 registered stations) and contains information about the station coordinates, instruments, operating networks, and when a station began/stopped operating. To facilitate data exchange between networks and organizations, it is strongly recommended that network operators register their stations and use the international station codes.

Summary

The high-quality data recorded by seismographic networks would be of limited value without international cooperation by all countries worldwide in the acquisition and exchange of seismic measurements and waveforms. This cooperation is essential for the location of earthquakes, for understanding the physics of earthquakes, and for studies of Earth's internal structure, properties, and processes. One of the latest challenges to acquiring, processing, and distributing data from seismographic networks by data centers globally is the automatic performance of traditional tasks in as close to real time as current technology permits.

Bibliography

Godey, S., Bossu, R., Gilbert, J., and Mazet-Roux, G., 2006. The Euro-Mediterranean Bulletin: a comprehensive seismological bulletin at regional scale. *Seismological Research Letters*, **77**, 460–474.

Godey, S., Mazet-Roux, G., Bossu, R., Merrer, S., and Guilbert, J., 2009. Ten years of seismicity in the Euro-Mediterranean region: Panorama of the EMSC Bulletin 1998–2007 (http://www.emsc-csem.org/docs/publications/ECGS_2009.pdf), Cahiers du Centre Européen de Géodynamique et de Séismologie, Vol. 28, pp. 1–14.

Liu, R.-F., Gao, J-Ch, Chen, Y.-T., Wu, Z.-L., Huang, Z.-B., Xu, Z.-G., and Sun, L., 2008. Construction and development of China digital seismological observation network. *Acta Seismologica Sinica*, **21**, 535–541.

Cross-references

Earthquake, Aftershocks
Earthquake, Location Techniques
Earthquake, Magnitude
Seismic Monitoring of Nuclear Explosions
Seismology, Monitoring of CTBT

SEISMOLOGY, GLOBAL EARTHQUAKE MODEL

Peter Suhadolc
Department of Geosciences, University of Trieste, Trieste, Italy

Definition

Exposure: Elements at risk; an inventory of those people or artifacts (and their characteristics) that are exposed to a hazard.

Seismic Hazard: A potentially damaging physical event, a phenomenon induced by an earthquake that may cause loss of life or injury, property damage, social and economic disruption, or environmental degradation. Seismic hazards can be single, sequential, or combined in their origin and effects. Each hazard is characterized by its location, intensity, frequency, and probability.

Seismic Risk: The probability of harmful consequences, or loss of life or injury, property damage, social and economic disruption or environmental degradation, resulting from interactions between seismic hazards and vulnerable conditions.

Vulnerability: The degree of loss of life or injury, property damage, social and economic disruption, or environmental degradation to a given element (or set of elements) resulting from a given hazard at a given level of intensity.

Introduction

A Global Earthquake Model (GEM) is being developed to establish an independent, uniform standard to calculate and communicate earthquake risk worldwide. By developing much-needed tools and software for reliable seismic risk assessment for basic and expert users in all regions of the world, GEM will provide necessary input for increased earthquake risk awareness and the undertaking of mitigating action.

The GEM initiative is organized as a public–private partnership, while an international community of scientists and professionals drives the development of the global earthquake model.

The work for GEM started in 2009, and a first comprehensive model plus accompanying tools and software, will become available by the end of 2013. Although much of the work is under development at the moment this article is being written, it is felt to be important to report on the initiative in this encyclopedia, because the global earthquake model will constitute an important contribution to the scientific community.

Scientific needs for GEM

The assessment and subsequent mitigation of earthquake risks is among the ultimate goals of both applied seismology and earthquake engineering. In spite of the fact that earthquake occurrence is globally a steady process and that most seismic events occur in uninhabited areas, the explosion in urban development and the enormous growth in the number of megacities in earthquake-prone areas (Figure 1), has considerably increased the seismic risk worldwide. Both cities in developed countries, like Tokyo or Los Angeles, and cities in developing countries, like Kathmandu or Jakarta, can suffer substantial damage due to a large earthquake in the next future. In fact, over 600,000 people died in the last century due to earthquakes (reference to USGS web site). Most of those deaths occurred in developing and emerging countries.

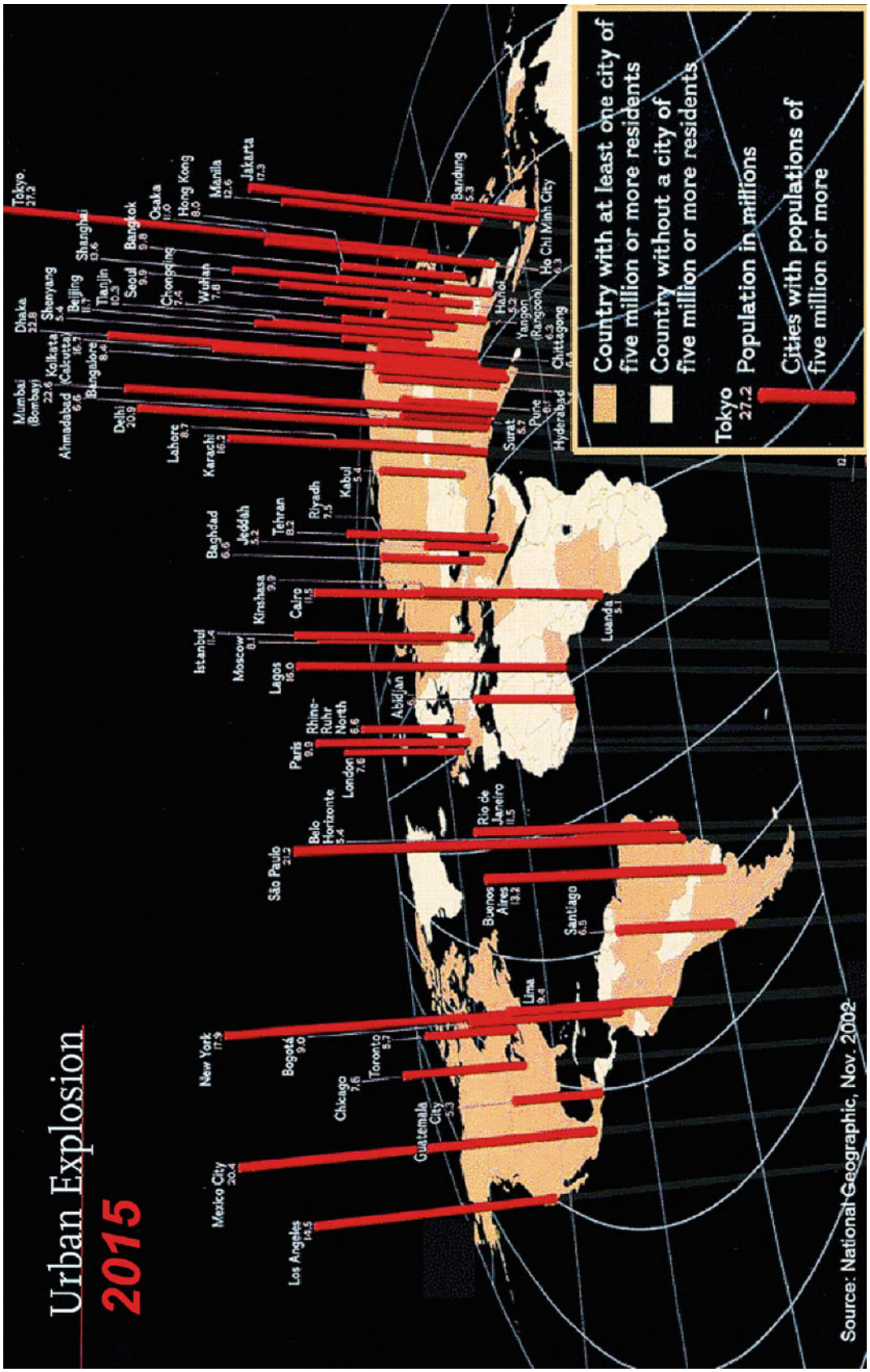
In many earthquake-prone regions, no seismic risk models exist, and even where they exist, they are not widely accessible. Such models are needed for accurate assessment of risks by (local) scientists, engineers, and practitioners, in order to promote mitigating actions, such as improvement of building codes and construction, sustainable land use, improved response, and protection of critical infrastructures. In order to have an effect on society at large, however, such models and the information resulting from it are needed as well by individuals and (international) organizations, to become aware of seismic risk and to undertake mitigating actions such as obtaining insurance, improved project management, and allocation of budgets for hazard mitigation.

The GEM initiative

By functioning as a community effort, the GEM initiative will produce a state-of-the-art dynamic and updatable model for the assessment of seismic risk worldwide; a model that is based on the probabilistic assessment of earthquake occurrence, the resulting ground motions, and the impact these have on structures and populations in terms of damage, social and economic loss. A model with underlying databases, that can be improved and enlarged with future data, and can be openly accessed through user-friendly software for data analysis and production of results, that can be also improved as our knowledge and technical capabilities rise in the future.

The global earthquake model is being designed and built by hundreds of experts and practitioners around the world, to ensure that less-monitored areas are also covered and to establish uniform standard, which allow for risk comparisons between countries and regions and for benchmarking output obtained through other sources. The model will reflect the needs, knowledge, and data of a variety of end users through GEM's extensive partner network. Such partnerships are essential in making sure that the information reaches the people that need it.

Technology transfer on the use of the software together with workshop opportunities will be provided, especially



Seismology, Global Earthquake Model, Figure 1 Urban explosion; it is predicted that by 2015 the projected increase in global population of 3 billion will concentrate itself in urban areas. (From National Geographic, 2002.)

in areas where risk assessment tools and data are currently less available.

A cooperative public–private partnership

The construction of the Global Earthquake Model is a cooperative public–private endeavor, and GEM is, therefore, structured as a partnership among country governments, private organizations, research institutions, international organizations and global risk initiatives, NGOs and individuals. The partnership includes a number of authoritative global institutions, such as the World Bank, the OECD, UNESCO, and UN's International Strategy for Disaster Reduction, but also the two largest international professional associations in the field: IASPEI (International Association of Seismology and Physics of the Earth's Interior) and IAEE (International Association for Earthquake Engineering), countries such as New Zealand, Norway, and Switzerland, and a number of prominent private corporations. Moreover, there are hundreds of institutions, organizations, and individuals involved in GEM that contribute expertise, data, or software, respond to requests for proposals, participate in regional programs, and take part in reviews and public assessments. Participation of individuals and institutions worldwide ensures that the model is owned by the global community and reflects its needs and knowledge.

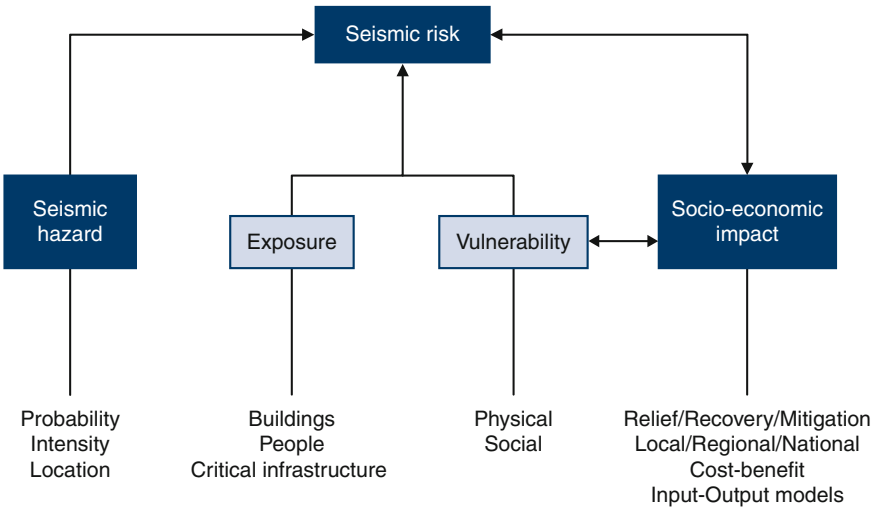
History

The idea for a Global Earthquake Model was born during a workshop called "Earthquake Science and its Contribution to Society" organized by the OECD's Global Science Forum in 2006. It was felt that a Global Earthquake Risk Scenario Map should be created. Various workshops followed in which the idea transformed into the creation of a Global Earthquake Model. Munich Re decided to sponsor the initiative in 2007. In 2008, the business plan for GEM was finalized and more partners were brought in to create a public–private partnership. The Eucentre was awarded the GEM Secretariat at the end of that year. In March 2009, the GEM Foundation was incorporated as a nonprofit foundation and that marked the start of the GEM initiative.

Scientific framework

The GEM scientific framework serves as the underlying basis for constructing the model, and consists of three principal integrated modules (Figure 2: Scientific Framework): Hazard, Risk, and Socioeconomic Impact.

Seismic risk is defined as a product of seismic hazard (the probability of levels of ground shaking, resulting from earthquakes, within a given time span), seismic vulnerability (the probability of loss given a level of ground shaking), and exposure (the elements at risk – mainly buildings, critical infrastructure, and humans). Risk gives an indication of the extent of loss (damage, fatalities, casualties) that can be expected in a given location in the world. Risk can, therefore, be high in an area without significant probabilities of ground shaking, because it has an older, more vulnerable, and densely populated building stock, and lower in an area with high levels of seismicity, but with well-constructed structures that are sparsely inhabited. Earthquakes, however, have an impact that goes beyond physical damage or casualties. Earthquakes can severely damage the economy and influence society and social well-being. Therefore, GEM will include innovative methods for analysis and evaluation of the impacts of earthquakes on the short, medium and long term, on local and global scales. There will also be applications that build upon the model, such as a tool for cost-benefit analysis, allowing users to understand what effect certain mitigation actions, strengthening of the building stock for example, will have on the risk. Insight into earthquake effects over time will directly support decisions on



Seismology, Global Earthquake Model, Figure 2 The current design of the scientific framework of GEM.

short-term needs after an event (relief, shelter), medium-term needs (recovery and reconstruction), and long-term needs related to policies on risk mitigation.

Implementation and current status

In June 2010, the GEM initiative has been able to deliver a proof-of-concept for hazard and risk calculations on a global scale. This will be the basis for development of the final GEM risk engine and the model. The model building process and engine are described in more detail further on.

International consortia, involving hundreds of professionals and institutions, are working on the creation of necessary standards, databases, and methodologies on a global level. These are the global components of the model. The work on Hazard Global Components has started and will be delivered in 2012. The work on Risk Global Components will start in the fall of 2010 and will be delivered in 2012 and 2013 and the work on the Socio-economic Global Components will take off in early 2011, with the goal to be finalized in 2013. The Global Components are further specified below.

Programs are being setup in many regions of the world as independently run, bottom-up projects, and links are established with ongoing regional programs. Both such programs are defined as GEM Regional Programs and involve a great number of local experts who will use GEM software, will generate local data, will validate the data and standards that were created on a global level, and will serve as a starting point for technology transfer in the region. Currently, three GEM Regional Programs are operational: in the regions of Europe and the Middle East, and collaboration is ongoing in Central America. Programs are being prepared in Africa, South-Asia, South-East Asia and Oceania, Central Asia, South America, the Caribbean, North-East Asia.

Global components

Global Components are the scientific modules of GEM that are developed at a global scale to provide standards, models, tools and data. Global components will provide the basic input to the model, but the Regional Programs will deliver detailed feedback and input from a local point of view.

To ensure that Global Components are developed by the international scientific community, GEM releases Requests for Proposals. These RfPs are developed by a group of international experts on the topic, and peer reviewed through an open public commenting system. International consortia, groups of institutions, and individual experts respond to these calls. All proposals are peer-reviewed by at least four external reviewers and extensively discussed by GEM's Scientific Board, who prepares an advice for GEM's Governing Board. Finally, for each component, one consortium is selected to carry out the work.

At the moment of this writing, proposals for the hazard and risk global components have been selected and the work has started or is about to start. The global component for socioeconomic impact is under development.

The hazard module

Five global components on seismic hazard are being developed as input to the global earthquake model:

Global historical catalog and database

The record of past earthquakes is among the most important means to evaluate earthquake hazard, and the distribution of damage associated with past earthquakes is a key to assessment of seismic risk. The instrumental seismic record has only a 100-year span, and yet no plate boundary is so active that this period is sufficient to capture the full range of seismic behavior. Extending the record of large damaging earthquakes several hundred years longer, and in exceptional cases by 1,000 years, is thus extremely valuable. This requires damage descriptions to be converted to numerical intensity scales, and the estimation of magnitude, location, and their uncertainties. Currently, an International consortia is working on the best and most efficient way to make use of the historical earthquake record, honoring at the same time its uncertainties and regional differences in quality and extent.

Global instrumental seismic catalog

An International consortium is building a reliable and uniform global instrumental earthquake database (1900–2009). A uniform location procedure to relocate global earthquakes from 1900 to 2009 shall be used, and standardized methods for computing both associated homogeneous surface-wave magnitude ($M_{S(\mathrm{BB})}$) using amplitude and period data and associated moment magnitude (M_W) from published seismic moments (M_0) and also proxy values, $M_{W[\mathrm{Mx}]}$, converted from other types of primary magnitudes [$\mathrm{M_x}$] using empirical relationships (Scordilis, 2006; Bormann et al., 2009; Bormann and Yadav, 2010). A recent example of such catalog structure and philosophy has been published by Yadav et al. (2009). In addition to producing a uniform global earthquake catalog, the database will also keep track of original input data files and documentation. These materials will allow seismologists to extend earthquake studies to a lower magnitude threshold for a better coverage of seismicity, especially in local and regional areas. The catalog will be the primary tool to be used to characterize the spatial distribution of seismicity, the magnitude–frequency relation, and the maximum magnitude.

Global active fault and seismic source database

Tectonic earthquakes are fault ruptures; seismic hazard assessments should, therefore, incorporate an inventory of active faults. Despite this, many seismic hazard assessments do not consider faults at all, or do so only sparingly because the requisite fault data are absent or inadequate. The need to incorporate active faults in the computation of a seismic hazard map, in order to estimate reliably the future strong ground motion, has been recognized long ago (Wesnousky et al., 1984; Wesnousky, 1986). Within the scope of GEM, a uniform global active fault and seismic source database is built, with a common set of

strategies, standards, and formats. The database will be publicly available. The effort is building upon the 1989–2004 International Lithosphere Program's Major Active Faults of the World. Since some faults cut the earth's surface and others do not, there will be both observational elements (active faults and folds that have slipped in the past 10–100 kyr) and interpretative elements (inferred seismic sources) to the database. In addition to collecting the best fault information presently available, appropriate mechanisms to capture new fault data as it becomes available, to capture the rapid expansion of fault knowledge and the ongoing debate about fault geometry, kinematics, and interaction are also being developed.

Global geodetic strain rate model

In the assessment of seismic hazard, seismic catalogs and active faults database provide complementary means for identifying zones of earthquake hazard. The geodetically measured secular strain rate provides an independent benchmark for crustal deformation and thus the recurrence of large earthquakes. The creation of a comprehensive and uniform model for geodetic strain rates is the scope of this global component. The consortium will critically review all global and regional studies since 1994 and will significantly update the Global Strain Rate Model of 2004. Moreover, they will assess whether the estimated fault slip rates and earthquake activity rates are consistent with the long-term strain accumulation measured from Global Positioning System (GPS) (or derived from interferometric synthetic aperture radar – InSAR).

Global ground-motion prediction equations (GMPEs)

A source of variability in hazard results obtained in small regions has traditionally been the derivation and use of local GMPEs, which are often based on insufficient data. Another factor influencing the variability of hazard results is the large statistical uncertainty of ground motion predictions; in recent years, a number of promising advances have been made – the Next Generation Attenuation (NGA) models (Power et al., 2006) – allowing to correctly account for various effects such as local site conditions, style-of-faulting effect, hanging-wall effect, etc., on ground motion predictions. Within the scope of GEM, by compiling a global reference hazard assessment model, a harmonized suite of GMPEs is developed, built on the most recent advances in the field.

Local soil conditions have a large influence on the seismic ground motions. The International consortium involved in this component is furthermore working on adopting a unified strategy to (a) determine seismically effective soil parameters, preferably based on the average S-wave velocity over the upper 30 m, V_{S30}, or, alternatively, derived from topography data (Wald and Allen, 2007; Allen and Wald, 2009) on a worldwide level, and (b) to represent the spatial distribution of soil classification, as for example proposed by NEHRP (Martin, 1994; BSSC, 2004) or EuroCode8 (CEN, 2004), compatible with the GMPE models.

The risk module

Five Global Components on seismic risk are being developed as input to the global earthquake model:

GEM ontology and taxonomy

Within the fields of hazard and risk assessment, diverse terminology is used and different meanings are sometimes attached to the same words. In order to achieve a shared understanding across the disparate fields and endeavors encompassed by GEM, an international consortium will develop some methodological foundation and terminology. Ontology refers to the entire framework that guides the development of the global earthquake model – the set of concepts and the relationship between these concepts that will allow determination and communication of earthquakes risk. The GEM Ontology will be general and comprehensive enough to be long-lasting, but must be adaptable to future conditions. Taxonomy, a part of ontology, refers to the classification of things in an ordered system, reflecting their relationships. The GEM Taxonomy will be an earthquake-related system of classification, set of terminology, encompassing hazard, risk, and socioeconomic realms. The adopted Ontology and Taxonomy will be evaluated and tested and finally globally promoted and disseminated in a continuous way.

Global earthquake consequences database

The international consortium working on this global component will create a global database of earthquake impacts and consequences. They will assemble and store in a structured and web-accessible way both data (including photos) already acquired and data yet to be acquired following future events. Data (both statistical and observational) typically covers building damage, damage to lifelines and other infrastructures, ground failure, human casualties, social disruption, and financial and economic impacts. All damages will be geographically referenced and viewable on a global mapping system. The database will be equipped with analytical tools enabling data fields to be post-processed across events, globally or within regions.

Global exposure database

Compiling an open database of the global building stock distribution with associated uncertainties containing the spatial, structural, and occupancy-related information for damage, loss, and human casualty estimation models is the focus of this Global Component. The International Consortium working on this will identify, evaluate, and homogenate the various existing databases at country, regional, and city levels throughout the world. Within an appropriate grid cell resolution, the inventory will include the total number of buildings, their floor area, the relative distribution of building types (e.g., timber, masonry, R/C, steel) along with some performance-influencing features, such as construction quality and year of construction. Moreover, the relative distribution of occupancy types (e.g., residential, industrial or agricultural) and the

temporal population (e.g., day- and nighttime) within each cell and for each building type will be considered. Finally, the presence of critical facilities and infrastructures for emergency relief will also be identified for each cell.

Inventory data capture tools

This Global Component aims to provide the tools that will enable the capture and transfer of high-resolution inventory data into either the Global Exposure Database or Global Earthquake Consequences Database. Both data collected using remote sensing or acquired from direct observation will be merged in this process.

Interpretation of low, medium, or high spatial resolution satellite imagery and of aerial images is foreseen for data capturing. This will involve the determination of geo-referenced footprints of buildings, vegetation (important for post-event fire spread), and other infrastructure (e.g., roads, lines, reservoirs). Appropriate software will upload and aggregate such data into a neighborhood or grid level of resolution. Direct observation on the other hand implies street-front direct inspection or interpretation of street-front data from photographs.

Global vulnerability estimation methods

The global vulnerability assessment comprises the estimation of the degree of direct loss due to building damage. Direct loss means both human losses (injuries and fatalities) as well as economic losses (repair costs, downtime, etc.) arising from the damage due to ground shaking at a given level. The goal of this global component is to provide methods and standards for vulnerability assessment that can be applied to a wide taxonomy of structures. The International consortium working on the component will derive a number of initial "default" functions that can be applied on a global scale. The methods will account for varying levels of detail in the input data, include the characterization of uncertainties, and account for the influence of retrofitting on the vulnerability.

The socioeconomic impact module

This module is considered as a single global component, which will allow for assessment of social and economic consequences of earthquakes, further and beyond those direct losses considered in the risk module, with the goal to provide users with a wide array of methods useful to communicate impacts of seismic events. The module is envisioned as a toolbox that shall gather a comprehensive set of models, metrics, data, and tools to be organized following spatial, temporal, and user principles under a standardized, clear, and simple framework.

To achieve this goal, GEM will propitiate the development of a framework for compiling state-of-the-art, widely accepted, useful methods for diagnostics and decision-making. The toolbox will be initially populated with methods generated by GEM in partnership with an international consortium, and after that through a constant process of interactions and consulting with the wider community. The toolbox will permit, to the extent possible, the integration of methods, and include interfaces for end users to input data and parameters, to conduct sensitivity and counterfactual analysis, to evaluate alternative policy interventions, and, to visualize results. Case studies are expected to be generated using the methods included in the toolbox.

All the tasks to be conducted within the social and economic impact module shall be performed following a participatory framework that will encompass GEM, its partners, and the wider community. This must be the case, given the multidimensional character and the complexities involved in linking variables from the natural, social, and economic systems, and considering that no consensus has been perceived regarding a precise definition of social and economic impacts of earthquakes and the methods for measuring or addressing them.

Model building

Currently, GEM risk engine is being developed. The engine will allow for calculations on a global level and, therefore, needs to be able to incorporate the data, standards, models, and methods developed by the international community through the Global Components and Regional Programs. A proof-of-concept of the engine has been delivered in June 2010, demonstrated by first hazard and risk calculations on a global scale, which resulted in preliminary global output. This output, however, has not been validated and will only be used internally to propel further development. The GEM risk engine is characterized by open-source development; hence development and enhancement by a wider community of expert users and programmers. It will be platform-independent, modular (using object-oriented language), flexible (as to allow for future multi-hazard calculations), expandable (in terms of methodologies employed), and scalable.

In order to serve the needs of various users, an understandable user interface is being constructed encompassing tools and software for transparent earthquake risk calculations, and risk communication, incorporating the latest technologies for sharing and contributing data between users.

Outcome and future

GEM is going through a continual user-needs assessment effort, to ensure that the software and tools that are being developed meet the needs of a wide range of possible users. Partnerships and an active user-community are the ingredients that support the initial use of the tools and subsequent adoption of the information that the global earthquake model produces, as a necessary first step toward awareness and risk-mitigating behavior.

The main output of GEM's first 5-year working program will be the OpenGEM platform, which will allow basic and expert users to run applications, access seismic risk information on local, national and regional scale, and visualize the latter in maps, curves, tables and export these in compatible formats. Basic users are likely to want

to view output produced by the global earthquake model, perhaps that related to the location of their own house. Expert users will be able to "plug in" their own data and run their own calculations. Because not everyone will be able to access an internet portal, or would like to run calculations through the internet, a stand-alone OpenGEM software package will be an important derivative.

GEM will, however, produce more than a platform for risk assessment. Global harmonized databases within the fields of earthquake hazard, vulnerability, exposure, and socioeconomic impact will be made available, such as a global earthquake consequences database and a global historical seismic catalog. GEM will also produce best practices and standards related to many aspects of seismic risk assessment, which will help the community to work together under a common framework at a global scale. A community development platform for the computational engine will allow for open-source and object-oriented development of the GEM risk engine by the community. Programmers and other experts will be able to test, use, and further improve GEM's software code. Finally, there will be technical reports for the (scientific) community to use, and technical training programs and workshops for diffusion of the knowledge on GEM software and use.

After completion of the first working program, GEM will continue its activities and work on extending the model, maintenance and improvement of the tools, will support the needs of an ever-growing user community, will work on dissemination of products and results and the development of new applications and partnerships.

The future challenge for the scientists will probably be not so much to invalidate the concept of GEM, but to improve its global components (methods, data and standards), as our scientific knowledge of the earthquake generation, seismic wave propagation, and their impact on the built environment – with all the related consequences – will advance in the years to come.

Current developments of the GEM initiative can be followed on www.globalquakemodel.org.

Acknowledgments

I would like to thank Nicole Keller, the GEM Outreach and Communications Officer, for the help she has given me and for providing Figure 1. Thanks also to David Wald for his constructive criticism on the preliminary version of this contribution.

Bibliography

Allen, T. I., and Wald, D. J., 2009. On the use of high-resolution topographic data as a proxy for seismic site conditions (V_{S30}). *Bulletin of the Seismological Society of America*, **99**, 935–943.

Bormann, P., and Yadav, R. B. S., 2010. Reply to comment on a homogeneous and complete earthquake catalog for Northeast India and the adjoining region by R.B.S. Yadav, P. Bormann, B. K. Rastogi, M.C. Das, and S. Chopra by R. Das and H.R. Wason. *Seismological Research Letters*, **81**(2), 235–240.

Bormann, P., Liu, R. F., Xu, Z. G., Ren, K. X., Zhang, L. W., and Wendt, S., 2009. First application of the new IASPEI teleseismic magnitude standards to data of the China National Seismographic Network. *Bulletin of the Seismological Society of America*, **99**, 1868–1891.

BSSC, 2004. *NEHRP (National Earthquake Hazard Reduction Program) Recommended Provisions for New Buildings and Other Structures (FEMA 450), 2003 Edition*. Washington, DC: Building Seismic Safety Council, National Institute of Building Sciences.

CEN (Comité Européen de Normalisation), 2004. Eurocode 8: Design of Structures for Earthquake Resistance, Part 1: General rules, seismic actions and rules for buildings. European Standard EN 1998-1:2004. CEN: Brussels, Belgium.

Martin, G. M. (ed.), 1994. *Proceedings of the NCEER/SEAOC/BSSC Workshop on Site Response during Earthquakes and Seismic Code Provisions*. University of California, Los Angeles, November 18–20, 1992

Power, M., Chiou, B., Abrahamson, N., and Roblee, C., 2006. The Next Generation of Ground Motion Attenuation Models (NGA) project: an overview. In: *Proceedings of the 8th U.S. National Conference on Earthquake Engineering*, Paper No. 2022.

Scordilis, E. M., 2006. Empirical global relations converting Ms and mb to moment magnitude. *Journal of Seismology*, **10**, 225–236.

Wald, D. J., and Allen, T. I., 2007. Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, **97**, 1379–1395.

Wesnousky, S. G., Scholz, C. H., Shimazaki, K., and Matsuda, T., 1984. Integration of geological and seismological data for the analysis of seismic hazard: a case study in Japan. *Bulletin of the Seismological Society of America*, **74**, 687–708.

Wesnousky, S. G., 1986. Earthquakes, quaternary faults, and seismic hazard in California. *Journal of Geophysical Research*, **91**, 12587–12631.

Yadav, R. B. S., Bormann, P., Rastogi, B. K., Das, M. C., and Chopra, S., 2009. A homogeneous and complete earthquake catalog for Northeast India and the adjoining region. *Seismological Research Letters*, **80**, 609–627.

References to websites

USGS website: http://earthquake.usgs.gov/regional/world/world_deaths.php

Cross-references

Seismic Hazard

SEISMOLOGY, MONITORING OF CTBT

Scientific and Technical Advances in Seismology and their Relevance

Wu Zhongliang[1], Paul G. Richards[2]
[1]Institute of Geophysics, China Earthquake Administration, Beijing, China
[2]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

Introduction

Monitoring of the *Comprehensive Nuclear-Test-Ban Treaty* (CTBT) entails the detection, identification, and

characterization of atmospheric, underwater, and specifically underground nuclear tests, and the discrimination of nuclear explosions from other artificial and/or natural events such as quarry blasts and earthquakes (CTBTO PrepComm., 2009a). Matching the needs of verification deals with seismic, infrasound, and hydro-acoustic, as well as radionuclide monitoring systems. At present the CTBT verification system specified by the treaty has three major components: the International Monitoring System (IMS) with a global network of 337 monitoring facilities; the International Data Centre (IDC) for the processing of observational data; and an On-Site Inspection (OSI) regime that utilizes a series of high-resolution technologies and has many similarities to the emergency on-site observation of disaster areas such as the meizo-seismal regions of large earthquakes. In this verification system, seismic monitoring has been the leading subject, due to three aspects: (1) Since 1980 up to this time of writing (2010), all nuclear tests have been conducted underground, and seismic monitoring is the most efficient way to monitor tests in that environment. (2) Development of seismic monitoring played the leading role in the development of IMS. Compared to other monitoring techniques, seismic monitoring data, characterized by the real-time transmission of broadband/high-frequency continuous waveforms, plays a major role in the whole monitoring system. (3) Important concepts in CTBT monitoring, including evasion scenarios such as decoupling and mine-masking which are most directly a challenge to seismic monitoring, were directly or indirectly developed from seismic monitoring practice.

Entry *seismic monitoring of nuclear explosions* describes in detail the basic properties of explosion and earthquake signals and seismic technologies for the detection and identification of explosions. The main tools for such a monitoring purpose, local, regional, and global seismic networks, are introduced in the entry *seismological networks*. Sections following focus on recent scientific and technical advances in seismology and their relevance to CTBT monitoring.

Evaluation and design of the seismic monitoring system: a systems engineering perspective

Monitoring and verification require an international observational system composed of the following components: observational stations, communication links from the stations to data center/s, and tools for the analysis of signals recorded for verification purposes. The evaluation and design of the seismic monitoring system from the perspective of systems engineering is one of the significant advances in recent years. It is useful not only for the functioning and sustainability of the CTBT monitoring system but also for the operation of modern seismological networks, which not only record earthquakes but also have missions such as the fast seismological information service and the managing of huge amount of observational data. In 2008, the Project "International Scientific Studies of the Implementation of the CTBT Verification System" (ISS) was organized by the Preparatory Commission of CTBT Organization (CTBTO PrepComm., 2009b), in which *system performance evaluation* provided useful tools for the evaluation and design of the seismic monitoring system.

Evaluation is to assess the readiness and capability of the monitoring and verification system, while design is to plan the road map to ensure the expected readiness and capability. Design and evaluation of a seismological monitoring system includes four perspectives: (1) *physical perspective,* including primary seismic stations and arrays, auxiliary stations, national data centers, international data center, network maintenance centers, communication links, system for the automatic data processing and review of analyst, database, and data sharing devices; (2) *functional perspective,* including data transmission, automatic data processing, automatic seismic phase picking and event location, interactive phase picking and location by analysts, array processing for location, magnitude determination, determination of earthquake parameters such as moment tensor and radiated energy, discrimination tools, and tools for routine evaluation of the state-of-health; (3) *operational perspective,* including the design, implementation, management, and upgrading of the whole system; and (4) *scientific perspective,* dealing with Earth structure model, travel-time database, algorithm for location and phase picking, attenuation model and site correction for magnitude determination, Green functions or attenuation relations for source parameter determination, azimuth correction database for array detection, and event database for discrimination. It may be unnecessary for a seismologist to directly use the concepts and tools of systems engineering. However, methodologies such as specification of the system and its components as well as the relationships among the components, specification of the functions of the system and decomposition of system functional requirements into component functional requirements, and development of measures of performance at system level and component level, respectively, provide a clear concept for the systematic evaluation and design of the seismic monitoring system. This becomes especially important when dealing with a modern seismological monitoring system with many stations, working continuously in a real-time mode, being a technically complex system.

Advances in seismology, potentially relevant to CTBT monitoring

Recent decades have witnessed the fast development of seismology (*Seismological Grand Challenges* Writing Group, 2008). These advances have apparent relevance to CTBT monitoring. Location of earthquake epicenters using conventional methods has the uncertainty of up to 10^1–10^2 km. This uncertainty is significantly reduced by the modern methods of earthquake location using waveform cross-correlation and relative location technique (Richards et al., 2006). Recent investigation reveals that "repeating earthquakes," identified by waveform

cross-correlation, seem to be much more common than expected in seismic activity (e.g., Schaff and Richards, 2004), providing the evaluation of location accuracy with an innovative and efficient tool (e.g., Jiang and Wu, 2006). By waveform comparison and relative location, it becomes possible for seismologists to determine the accurate location of new tests if data from previous nearby tests is available (e.g., Wen and Long, 2010). As archives grow, new seismic events that appear problematic often can best be resolved by comparison of their signals with those of previously recorded events from the same region.

In the evaluation of location accuracy, one of the key concepts is the *ground truth* (GT) *events* (International Seismological Centre, 2009), a database of reference earthquakes and/or explosions for which the epicenter information is known with high confidence (to ±5 km or better) with seismic signals recorded at regional and/or teleseismic distances. GT events are used as "calibrating information" to assess the location accuracy of the monitoring system. Furthermore, a GT database is useful to facilitate better visualization of Earth structure, better modeling of velocities of seismic waves, and more accurate travel-time determinations, which are all essential to the enhancement of monitoring capability. Using and developing the GT event database is thus endorsed by international seismological organizations such as the International Association of Seismology and Physics of the Earth's Interior (IASPEI, 2009a).

Recent developments of seismology kept challenging some of the conventional ideas well-accepted in the CTBT monitoring practice. At present, there are several complementary methods for the identification of an explosion from an earthquake, such as the difference between m_b and M_S, and the ratio of high-frequency (>2 Hz) P to S energy (Richards and Zavales, 1990; Bowers and Selby, 2009). One of the important measures is the minimum size of earthquakes and/or explosions for the discrimination, called the *magnitude threshold,* below which the discrimination tool fails to work. Conventionally, the *magnitude threshold* for discrimination is regarded as different from, often 0.5 magnitude unit larger than, the *magnitude threshold* for detection (Hannon, 1985). However, new results (CTBTO PrepComm., 2009b) showed that these thresholds appear to be very similar if regional data, adequate for measurement of spectral ratios, is available. For purposes of estimating the location and yield of underground explosions, it is always important to have accurate information on the inhomogeneous structure of the Earth within which seismic waves propagate. New approaches to this problem using seismic noise (e.g., Shapiro et al., 2005) provide an innovative tool for imaging the Earth structure. Increase of seismic stations and development of inversion techniques kept improving the Earth structure model (*Seismological Grand Challenges* Writing Group, 2008). Accordingly, traditionally problematic earthquake parameters such as focal depth, and "modern" earthquake parameters such as moment tensor and radiated energy potentially can contribute more to CTBT monitoring, along with improved 3D Earth models, and accurate account of wave propagation in complex 3D media.

With the development of digital seismic waveform analysis techniques, it has also become possible for seismologists to capture the time-lapse behavior of Earth medium with considerable accuracy (e.g., Grêt and Snieder, 2005). Combining with the recent advances in active source technology, this development in turn has direct impact on the OSI technology. Combination of seismic data with remote sensing images shows potential in revealing more details of the explosion source, such as decoupling (Sykes et al., 1993). This combination, enabled by modern information technology such as "Digital Earth" (Wu and Chen, 2000), or more practically Google Earth®, provides an alternative approach that supplements the traditional concepts of OSI and even verification.

"Forensic seismology": evidences and judgments

Seismology is a branch of observational science based on the analysis of signals from different sources, penetrating through the Earth, and at last recorded by seismic networks. The field related to seismological observatory practice (Bormann, 2002) and the operation of seismic networks is sometimes called *seismological observation and interpretation* (IASPEI, 2009b). When seismology is applied to the monitoring of CTBT, more considerations are needed beyond the recording and analysis of seismic signals. Key concepts related to the practical functioning of the seismic monitoring system are *monitoring* and *verification*. The term *monitoring* refers to technical activities and analyses associated with data from observational systems that acquire signals from different sources, either explosions or earthquakes, in a continuous and, in some cases, real-time regime. The term *verification* refers to evaluations that include non-seismologists, to enable authorities and the public to judge whether some detected phenomena had its origin in a nuclear explosion. To this end, *verification* is regarded as a branch of *forensic seismology* (Bowers and Selby, 2009), in which the word "forensic" means answering questions of interest to (international) legal system/s.

In concept, *monitoring* is different from *verification* in that *monitoring* basically provides objective observational evidences, while *verification* has to have some subjective decision-making processes based on the objective (but in some cases limited) evidences provided by monitoring. An example is the announced test of DPRK in 2009. It was declared by DPRK authorities that the nuclear test was successfully conducted, but there were no direct evidences (radionuclide) to prove that the explosion, recorded clearly by seismic networks, was really a successful *nuclear* test. Another example is the 1998, Indian-Pakistan announced tests – even if seismic monitoring data could not provide persuasive evidences that *all* the tests were conducted as successfully as announced. In most cases, however, the monitoring system serves for

the clarification of special events, such as the earthquake near Lop Nor of China on March 13, 2003 (Bowers and Selby, 2009). In fact, an important task of the seismic monitoring system is to *screen* events of interest, to identify events that are definitely not nuclear tests. Note that in the context of forensic science, to verify that a nuclear test has been conducted somewhere and to verify that no nuclear test has been conducted in a certain place for a specific time duration, even when using the same monitoring system, need different sets of evidence from that system.

Concluding remarks: interaction between science and CTBT monitoring

The concept of seismic monitoring of nuclear tests can be traced back to the earliest underground nuclear explosions (Bolt, 1976; Richards and Kim, 2009). In September 1957, during the General Assembly of the International Union of Geodesy and Geophysics (IUGG), Toronto, K. E. Bullen gave the address "Seismology in our Atomic Age." Needs of nuclear test monitoring played an important role in the development of global seismic networks in the 1960s (Richards, 2002). In the meantime, researches have been carried out extensively for the detection and identification of nuclear tests (Husebye and Mykkeltveit, 1981; Richards, 2002), in which one of the remarkable techniques is seismic array for the detection of explosion signals smaller than those monitored by conventional stations (Douglas, 2002).

In recent years, the international CTBT monitoring system has been in operation as a "big science device" that is similar in scale to accelerators in high-energy physics and satellites in space science. It provides not only experiences and lessons in the practice of monitoring, but also unique datasets which are interesting in basic research and the science and technology for sustainability. Advances in science and technology also promote the implementation and development of CTBT monitoring practice, and assessments of the overall capability. In the above sections, these relevant advances are introduced via three components: *engineering*, *science*, and *decision making*. It is worth mentioning that, as indicated by a series of important events, especially the Conference "CTBT: Synergies with Science, 1996–2006 and Beyond" (CTBTO PrepComm., 2006) and the ISS Project (CTBTO PrepComm., 2009b), a new era of cooperation between CTBT monitoring communities and scientific research communities has started, which will in turn contribute both to the CTBT monitoring practice and to the development of geophysical science. In the ISS Project, *science and technology foresight*, a professional practice that emerged in recent years, aiming at identifying today's research and innovation priorities on the basis of scenarios of future developments in science and technology, society and economy, was introduced to the scientific assessment of CTBT monitoring, which indicates that the interaction between science and CTBT monitoring has started to be systematic and comprehensive through a professionally planned road map.

Bibliography

Bolt, B. A., 1976. *Nuclear Explosions and Earthquakes – The Parted Veil*. San Francisco: W. H. Freeman.

Bormann, P. (ed.), 2002. *IASPEI New Manual of Seismological Observatory Practice*. Potsdam: GFZ.

Bowers, D., and Selby, N. D., 2009. Forensic seismology and the comprehensive Nuclear-Test-Ban treaty. *Annual Review of Earth and Planetary Sciences*, **37**, 209–236.

CTBTO PrepComm. 2006. CTBT: Synergies with Science, 1996–2006 and Beyond. http://www.ctbto.org/the-organization/ctbt-synergies-with-science1996–2006-and-beyond/.

CTBTO PrepComm. 2009a. Verification Regime. http://www.ctbto.org/verification-regime/.

CTBTO PrepComm., 2009b. ISS 09 – International Scientific Studies. http://www.ctbto.org/specials/the-international-scientific-studies-project-iss/.

Douglas, A., 2002. Seismometer arrays – their use in earthquake and test ban seismology. In Lee, W. H. K., Kanamori, H., Jennings, P. C., and Kisslinger, C. (eds.), *International Handbook of Earthquake and Engineering Seismology*. Amsterdam: Academic, pp. 357–368.

Grêt, A., and Snieder, R., 2005. Monitoring rapid temporal change in a volcano with coda wave interferometry. *Geophysical Research Letters*, **32**, L06304, doi:10.1029/2004GL021143.

Hannon, W., 1985. Seismic verification of a comprehensive test ban. *Science*, **227**, 251–257.

Husebye, E. S., and Mykkeltveit, S. (eds.), 1981. *Identification of Seismic Sources – Earthquake or Underground Explosion*. Dordrecht: D. Reidel.

IASPEI, 2009a. Resolutions at the 2009 Scientific Assembly, Cape Town, South Africa. http://www.iaspei.org/resolutions/resolutions_2009_cape_town.pdf.

IASPEI, 2009b. Commission on Seismological Observation and Interpretation. http://www.iaspei.org/commissions/CSOI.html.

International Seismological Centre, 2009. Reference Event Bulletin. http://www.isc.ac.uk/GT/index.html.

Jiang, C. S., and Wu, Z. L., 2006. Location accuracy of the China national seismograph network estimated by repeating events. *Earthquake Research in China*, **20**, 67–74.

Richards, P. G., 2002. Seismological methods of monitoring compliance with the comprehensive nuclear test ban treaty. In Lee, W. H. K., Kanamori, H., Jennings, P. C., and Kisslinger, C. (eds.), *International Handbook of Earthquake and Engineering Seismology*. Amsterdam: Academic, pp. 369–382.

Richards, P. G., Waldhauser, F., Schaff, D. P., and Kim, W.-Y., 2006. The applicability of modern methods of earthquake location. *Pure and Applied Geophysics*, **163**, 351–372.

Richards, P. G., and Kim, W.-Y., 2009. Monitoring for nuclear explosions. *Scientific American*, **300**(3), 64–71.

Richards, P. G., and Zavales, J., 1990. Seismic discrimination of nuclear explosions. *Annual Review of Earth and Planetary Sciences*, **18**, 257–286.

Schaff, D. P., and Richards, P. G., 2004. Repeating seismic events in China. *Science*, **303**, 1176–1178.

Seismological Grand Challenges Writing Group, 2008. *Seismological Grand Challenges in Understanding Earth's Dynamic Systems, Report of the Long Range Science Plan for Seismology Workshop, September 18–19, 2008*. Washington: Incorporated Research Institutions for Seismology (IRIS). http://www.iris.edu/hq/lrsps/.

Shapiro, N. M., Campillo, M., Stehly, L., and Ritzwoller, M. H., 2005. High resolution surface wave tomography from ambient seismic noise. *Science*, **307**, 1615–1617.

Sykes, L. R., Deng, J., and Lyubomirskiy, P., 1993. Accurate location of nuclear explosions at Azgir, Kazakhstan, from satellite images and seismic data: implications for monitoring decoupled explosions. *Geophysical Research Letters*, **20**, 1919–1922.
Wen, L., and Long, H., 2010. High-precision location of North Korea's 2009 nuclear test. *Seismological Research Letters*, **81**, 26–29.
Wu, Z. L., and Chen, Y. T., 2000. Potential impact of "Digital Earth" on seismological monitoring. *Seismological Research Letters*, **71**, 548–552.

Cross-references

Seismic Monitoring of Nuclear Explosions
Seismological Networks

SEISMOLOGY, ROTATIONAL

William H. K. Lee
U.S. Geological Survey, Menlo Park, CA, USA

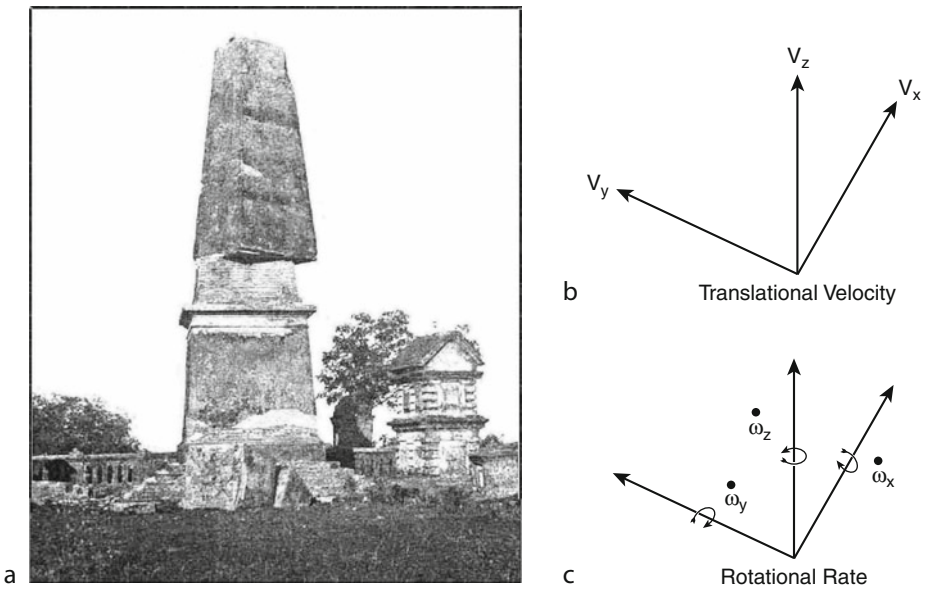
Definition

Rotational seismology is an emerging study of all aspects of rotational motions induced by earthquakes, explosions, and ambient vibrations. The subject is of interest to several disciplines, including seismology, earthquake engineering, geodesy, and earth-based detection of Einstein's gravitation waves.

Introduction

Rotational effects of earthquake waves together with rotations caused by soil-structure interaction have been observed for centuries (e.g., rotated chimneys, monuments, and tombstones relative to their supports). A summary of historical examples of observations on earthquake rotational effects is provided by Kozák (2009), including reproduction of the relevant sections from Mallet (1862) and Reid (1910). Figure 1a shows the rotation of the monument to George Inglis (erected in 1850 at Chatak, India) as observed by Oldham (1899) after the 1897 Great Shillong earthquake. This monument had the form of an obelisk rising over 60 ft. high from a base 12 ft. on each side. During the earthquake, the topmost six-foot section was broken off and fell to the south and the next nine-foot section was thrown to the east. The remnant is about 20 ft. in height and is rotated about 15° relative to the base.

A few early authors proposed rotational waves or at least some "vortical" motions. Many different terms were used for the rotational motion components at this early stage of the field's development. For example, "rocking" is rotation around a horizontal axis, sometimes also referred to as "tilt." Mallet (1862) proposed that rotations of a body on the Earth's surface are due to a sequence of different seismic phases emerging at different angles. Reid (1910) studied this phenomenon, which was observed in the 1906 San Francisco earthquake, and pointed out that the observed rotations are too large to be produced by waves of elastic distortion. Such waves "*produce very small rotations, whose maximum amount, . . . is given by the expression* $2\pi A = \lambda$*, where A is the amplitude and* λ *the wavelength; with a wave as short as 10,000 feet (3 km) and an amplitude as large as 0.2 of a foot (6 cm), the maximum rotation would only be about 0.25 of*



Seismology, Rotational, Figure 1 (**a**) Rotation of the monument to George Inglis (erected in 1850 at Chatak, India) as observed by Oldham (1899) after the 1897 Great Shillong earthquake. (**b**) Coordinate system for translational velocity. (**c**) Coordinate system for body-fixed or instantaneous rotational rate.

a minute of arc [0.0042 degrees, or 7.3 micro-radians (μrad)], *a quantity far too small to be noticeable*" (Reid, 1910, p. 44). A modern analysis of such rotational effects is presented in Todorovska and Trifunac (1990).

Observational seismology is based mainly on measuring *translational* motions because of a widespread belief that *rotational* motions are insignificant. For example, Richter (1958, footnote on p. 213) states that "Theory indicates, and observation confirms, that such rotations are negligible." Richter provided no references, and there were no instruments at that time sensitive enough to either confirm or refute this claim. Recent advances in rotational seismology become possible because sensitive rotational sensors have been developed in aeronautical and astronomical instrumentation. In this article we provide some background information and selected highlights of recent advances in rotational seismology. An extensive list of references to original papers is given because few "synthesis" articles exist.

Early attempts in studying rotational motions

Ferrari (2006) summarized two models of an electrical seismograph with sliding smoked paper, developed by P. Filippo Cecchi in 1876 to record three-component *translation* motions and also the *torsion* movements from earthquakes. Although these instruments operated for several years, no rotational motion could be recorded because of low transducer sensitivity. Pioneers in several countries attempted to measure rotational motions induced by earthquakes. Nearly a century ago, Galitzin (1912) suggested using two identical pendulums installed on different sides of the same axis of rotation for separate measurement of rotational and translational motion. This was later implemented, for example, by Kharin and Simonov (1969) in an instrument designed to record strong ground motion. Using an azimuthal array of seismographs, Droste and Teisseyre (1976) derived rotational seismograms for rock bursts in a nearby mine. Inspired by Walter Munk, Farrell (1969) constructed a gyroscopic seismometer, and obtained a static displacement of <1 cm and a tilt of <0.5 μrad at La Jolla, California, during the Borrego Mountain earthquake of April 9, 1968 (magnitude 6.5), at an epicentral distance of 115 km.

Early efforts also included studies of explosions. For example, Graizer (1991) recorded tilts and translational motions in the near field of two nuclear explosions using seismological observatory sensors to measure point rotations directly. Nigbor (1994) measured rotational and translational point ground motions directly with a commercial rotational MEMS sensor and found significant near-field rotational motions (660 μrad at 1 km distance) from a one-kiloton explosion.

Rotations and strains of the ground and of response of structures have been deduced indirectly from accelerometer arrays using methods valid for seismic waves having wavelengths that are long compared to the distances between sensors (e.g., Trifunac, 1979, 1982; Oliveira and Bolt, 1989; Spudich et al., 1995; Bodin et al., 1997; Huang, 2003, Suryanto et al., 2006; Wassermann et al., 2009). The rotational components of ground motion have also been estimated theoretically, using kinematic source models (Bouchon and Aki, 1982; Wang et al., 2009) and linear elastodynamic theory of wave propagation in elastic solids (Lee and Trifunac, 1985, 1987).

Measuring rotational motions

The general motion of the particles or a small volume in a solid body can be divided into three parts: translation (along the X-, Y-, and Z-axes), rotation (about the X- Y-, and Z-axes), and strain (six components). Figure 1b shows the axes in a Cartesian coordinate system for *translational velocity* measured by the usual seismometers in seismology, and Figure 1c shows the corresponding axes of *rotation rate* measured by rotational sensors (Evans, 2009). These are "body attached" coordinates, those that a seismic instrument would measure at a given instant as the sensors move and rotate through space. Converting an extended record of these body-fixed motions to recover motions in an Earth-fixed, quasi-inertial coordinate system has been performed for decades in "strapped down" inertial navigation systems such as those attached to a moving airplane. Lin et al. (2010) introduce these equations into seismology and earthquake engineering for recovering inertial-frame ground and structure motions.

Rotational ground motions can be measured directly by gyroscopic sensors or inferred indirectly from an array of translational sensors. According to Cochard et al. (2006), in a linear elastic medium the displacement $\mathbf{u}$ of a point $\mathbf{x}$ is related to a neighboring point $\mathbf{x} + \delta\mathbf{x}$ by

$$\mathbf{u}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{u}(\mathbf{x}) + \boldsymbol{\varepsilon}\,\delta\mathbf{x} + \omega \times \delta\mathbf{x} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the strain tensor and

$$\omega = {}^1\!/_2 \nabla \times \mathbf{u}(\mathbf{x}) \tag{2}$$

is a pseudo-vector representing the infinitesimal angle of rigid rotation generated by the disturbance. The three components of rotation about the X-axis, Y-axis, and Z-axis are given by the following equations for such infinitesimal motions:

$$\begin{aligned} \omega_x &= {}^1\!/_2\left(\partial u_z/\partial y - \partial u_y/\partial z\right), \\ \omega_y &= {}^1\!/_2\left(\partial u_x/\partial z - \partial u_z/\partial x\right), \\ \omega_z &= {}^1\!/_2\left(\partial u_y/\partial x - \partial u_x/\partial y\right) \end{aligned} \tag{3}$$

Therefore, rigid rotations can be observed by: (1) an array of translational seismometers indirectly for "cord" rotations associated with long wave lengths by assuming that contamination of translational signals by rotational motions is small, and that the linear elasticity theory is valid (e.g., Spudich and Fletcher, 2008), or (2) rotational sensors directly for "point" body-fixed rotations (e.g., Lee et al., 2009b; Lin et al., 2010).

In the past decade, rotational motions from small local earthquakes to large teleseisms were successfully recorded by sensitive rotational sensors in several countries (e.g., Takeo, 1998; McLeod et al., 1998; Igel et al., 2005, 2007; Suryanto et al., 2006; Cochard et al., 2006). In particular, the application of Sagnac interferometry in large ring laser gyros provided greatly improved sensitivity to rotations at teleseismic distances and showed that they are a good match to those estimated from linear elastic wave theory. Such motions provide additional observations that – combined with translations – lead to new approaches to the seismic inverse problem (Bernauer et al., 2009; Fichtner and Igel, 2009). Recently Kurrle et al. (2010b) reported the first observations of Earth's free oscillations using ring laser recordings, and opened up potential applications of rotational seismology at long periods.

In contrast, strong-motion observations near the source in both Japan and Taiwan showed that the amplitudes of these rotations can be one to two orders of magnitude greater than that expected from linear elasticity theory (e.g., Takeo, 1998; Lee et al., 2009b). Theoretical work suggests that in shallow granular or cracked continua (e.g., weathered rock at or near Earth's surface), asymmetries of the stress and strain fields can create rotations separate from and larger than those predicted by classical elastodynamic theory (e.g., Teisseyre et al., 2006, 2008). Unlike the traditional fault-slip model, Knopoff and Chen (2009) consider the case for faulting that takes place on a fault of finite thickness. They show that there is an additional single-couple term in the body-force equivalence and additional terms in the far-field displacement. They also show that the single-couple equivalent does not violate the principles of Newtonian mechanics because the torque imbalance in the single-couple is counterbalanced by rotations within the fault zone, with torque waves being radiated.

Large ring laser gyros

An unexpected advance in studying rotational ground motions came from a different field of geophysics. Recent developments of highly sensitive ring laser gyroscopes to monitor the Earth's rotation also yield valuable data on rotational motions from large teleseismic events. The most important property that makes such rotation sensors useful for seismology is its very low noise floor and high sensitivity to rotational motions and its insensitivity to translational and cross-rotational motions. The rotation rates expected and observed in seismology range from the order of 10^{-1} rad/s (e.g., Nigbor, 1994; Trifunac, 2009) near seismic sources down to order 10^{-11} rad/s for large earthquakes at teleseismic distances (e.g., Igel et al., 2005, 2007). This range spans at least 10 orders of magnitude, 200 dB, much as for translational motions, and it is unlikely that one instrument or one instrumental technology will be capable of providing accurate measurements over such a large range of amplitudes. Ring laser technology is currently the most promising approach to recording the small rotational motions induced by teleseisms, but the primary drawback is its very high cost.

Ring lasers detect the Sagnac beat frequency of two counter-propagating laser beams (Stedman, 1997; and Figure 2c). These active interferometers generally form triangular or square closed loops several meters across and are evacuated. If this instrument is rotating on a platform with respect to inertial space, the effective cavity length between co-rotating and counter-rotating laser cavity differs and one observes frequency splitting resulting in a beat frequency. This beat frequency δf is directly proportional to the rotation rate $\boldsymbol{\Omega}$ around the surface normal $\mathbf{n}$ of the ring laser system, as given by the Sagnac equation:

$$\delta f = \frac{4A}{\lambda P}\mathbf{n}\cdot\boldsymbol{\Omega}\,, \tag{4}$$

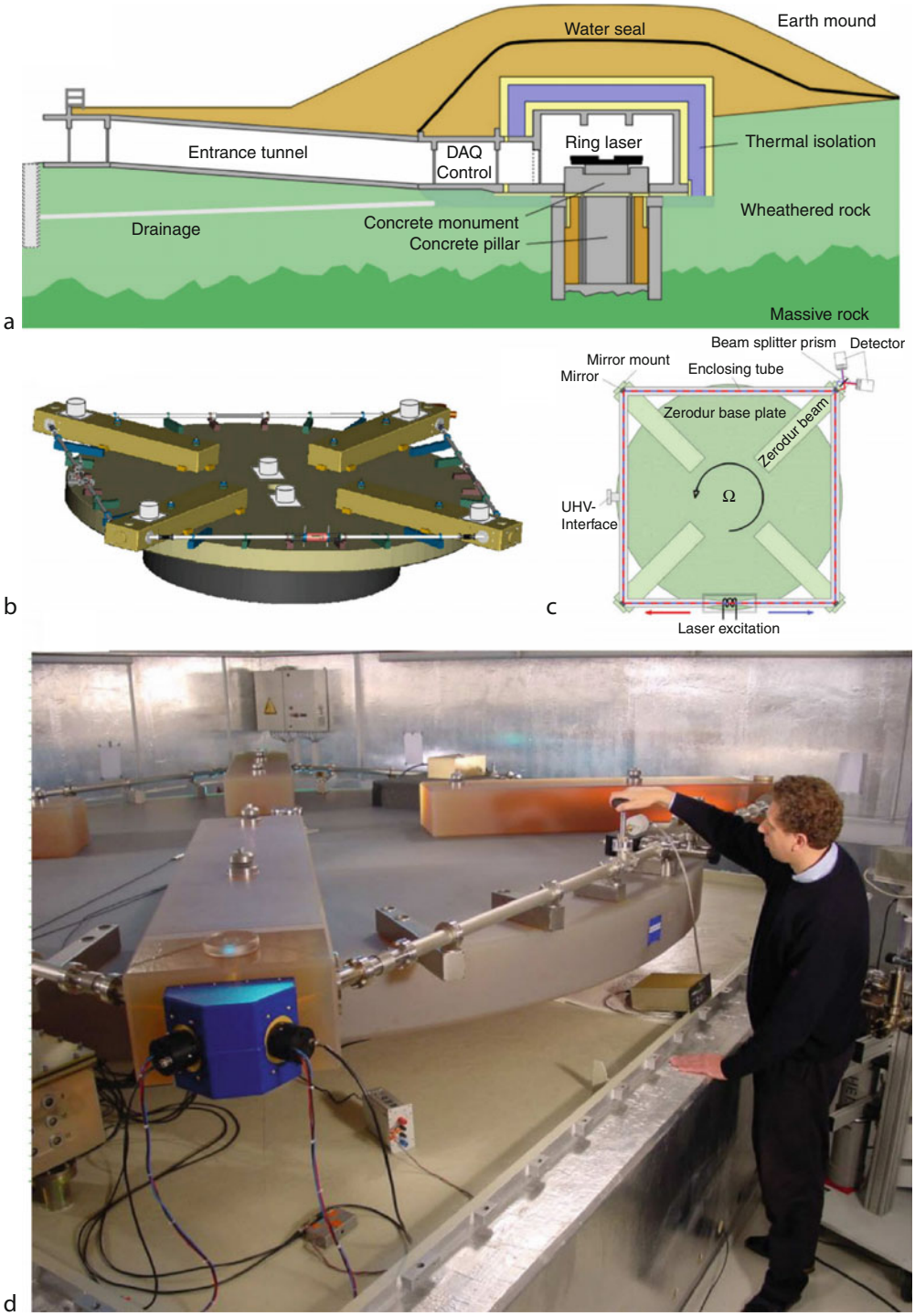
where P is the perimeter of the instrument, A its area, and λ the laser wavelength. This equation has three contributions that influence the beat frequency δf: (1) variations in the scale factor $(4A/\lambda P)$ have to be avoided by making the instrument mechanically as rigid and stable as possible, (2) changes in orientation $\mathbf{n}$ (tilting relative to Earth's rotation axis) enter the beat frequency via the inner product, and (3) variations in $\boldsymbol{\Omega}$ (e.g., changes in Earth's rotation rate and seismically induced rotations). Thankfully, the dominant contribution to δf is $\boldsymbol{\Omega}$. Note that translations do not contribute to the Sagnac frequency unless they affect P or A in some indirect manner.

Ring lasers are sensitive to rotations only, assuming stable ring geometry and lasing. However, for co-seismic observations at the Earth's surface the horizontal components of rotation (i.e., tilts) will contribute to the vertical component of rotation rate. As shown by Pham et al. (2009), the tilt-coupling effect is several orders of magnitude below the level of the earthquake-induced rotational signal unless one is very close to the source (where sensitive ring lasers would not be the appropriate technology).

At present, there are ring laser gyros capable of measuring rotation (induced by small local earthquakes or distant large teleseisms) at four sites: (1) Cashmere cavern, Christchurch, New Zealand (McLeod et al., 1998); (2) Wettzell, Germany (Schreiber et al., 2005); (3) Conway, Arkansas (Dunn et al., 2009); and (4) Piñon Flat, California (Schreiber et al., 2009a).

G Ring laser and recording teleseisms

Since 2001, the "G Ring" laser (capable of measuring rotation rate of about 10^{-10} rad/s) has been operating at the primary geodetic station (Fundamentalstation) at Wettzell, in Bavaria, (http://www.fs.wettzell.de/). A cross-sectional view of the site of the G Ring laser is shown in Figure 2a. The instrument is resting on a polished granite table (Figure 2b) embedded in a 90-t concrete monument. As shown in Figure 2a, the monument is



Seismology, Rotational, Figure 2 G Ring laser gyro at the Wettzell Superstation, Germany. (**a**) Cross-sectional view of the instrument site. (**b**) Instrument resting on a granite table. (**c**) Schematic drawing. (**d**) Photo of G Ring laser gyro with its designer, Ulli Schreiber.

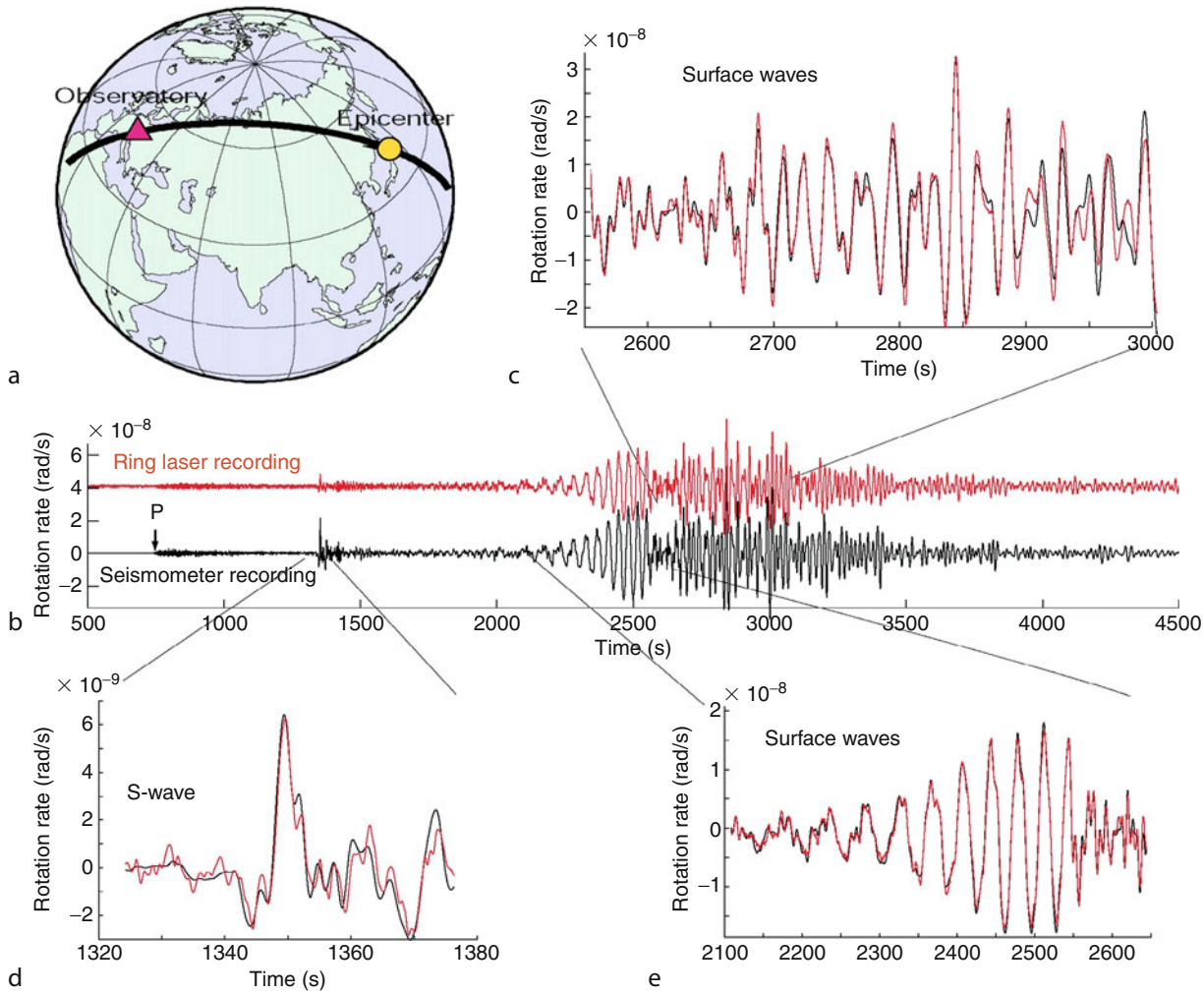
attached to a massive 2.7-m diameter concrete pillar and this is founded on crystalline bedrock 10 m below. A system of concrete rings and isolation material shields the monument and pillar from adjacent weathered rock to eliminate its deformation and heat-flow contributions. The G Ring laser is protected against external influences by a subsurface installation with passive thermal stability provided by a 2-m layer alternating between Styrofoam

and wet clay, this beneath a 4-m soil mound. A lateral entrance tunnel with five isolating doors and a separate control room minimize thermal perturbations during maintenance. After 2 years of thermal adaptation, the average temperature reached 12.2°C with seasonal variations of less than 0.6°C. Figure 2c shows the schematic drawing of instrument, and Figure 2d is a photo of the G Ring laser with its designer, Ulli Schreiber.

Figure 3 is a comparison of direct point measurements of ground rotations around a vertical axis (red lines) to transverse accelerations (black lines, converted to rotation rate for each time window) for the M8.1 Tokachi-oki earthquake, September 25, 2003 (Igel et al., 2005). Figure 3a is a schematic view of the great-circle-path through the epicenter in Hokkaido, Japan, and the observatory in Wettzell, Germany. 3b–e show the superposition of the rotation rate derived from transverse translations (black) and measured directly (red) for various time windows: Figure 3b the complete signal, Figure 3c the latter part of the surface wave train, Figure 3d the direct *S*-wave arrival, and Figure 3e for the initial part of the surface wave train. These results confirm the expectation from linear elasticity that the waveforms of transverse acceleration and rotation rate (around the vertical axis) should be identical assuming plane harmonic waves. Information on subsurface structure is contained in the ratio between the corresponding motion amplitudes.

Strong-motion inertial angular sensors

In aerospace, automotive, and mechanical engineering, smaller rotational-motion sensors are common and generically known as gyroscopic or inertial angular sensors. Nigbor (1994) used a MEMS-based Coriolis rotation-rate sensor to measure the rotational components of strong



Seismology, Rotational, Figure 3 Comparison of direct measurements of ground rotational motions around a vertical axis (*red lines*) with transverse accelerations (*black lines*, converted to rotation rate for each time window) for the M8.1 Tokachi-oki earthquake, September 25, 2003.

ground motions close to a large chemical explosion. Similar sensors were used by Takeo (1998) to measure rotational motions from an earthquake swarm ~3 km away. However, such sensors do not have the sensitivity to record rotations from small local earthquakes (magnitude ~4) at distances of tens of kilometers.

The eentec™ model R-1™ rotational seismometer is the first modestly priced sensor capable of recording small earthquakes at distances up to several tens of kilometers. It uses electrochemical technology in which the motion of an electrolytic fluid inside a torus is sensed electronically, yielding a voltage signal proportional to rotational velocity. Nigbor et al. (2009) carried out extensive tests of commercial rotational sensors and concluded that the R-1 sensor generally meets the specifications given by the manufacturer but that clip level and frequency response vary from those specifications and between individual channels enough that more detailed calibrations are warranted for each unit. A typical transfer function for the R-1 can be found at the manufacturer's website (http://www.eentec.com/). The instrument response is roughly "flat" from 0.1 to 20 Hz, and its self noise (rms) is <10 μrad/s over the same frequency band.

The R-1 rotational seismometers successfully recorded several hundred local earthquakes and two explosions in Taiwan (Lee et al., 2009b). Figure 4a shows the instruments deployed at station HGSD in eastern Taiwan (Liu et al., 2009). The top frame is a schematic drawing of the various seismic, geodetic, and strain instruments there. The bottom frame shows the subset of the instruments deployed in the shallow vault at the left hand side of the upper drawing; these include a datalogger (Quanterra Q330), an accelerometer (Kinemetrics Episensor), a six-channel digital accelerograph (Kinemetrics K2 with an external rotational seismometer, R-1 by eentec), and a short-period seismometer (Mark Products L-4A). The K2+R-1 instrument is at the left hand side, and the yellow-color box is the R-1 rotational seismometer.

The largest peak rotational rate recorded at the HGSD station (to early 2008) is from a M_w 5.1 earthquake at a hypocentral distance of 51 km at 13:40 UTC, 23 July 2007. Figure 4c shows the amplitudes and spectra of translational accelerations recorded by the K2's accelerometer. The peak ground acceleration was 0.47 m/s^2, and the two horizontal components have much higher amplitude than the vertical. Figure 4d shows the amplitudes and spectra for rotational rates recorded from its external R-1 seismometer. The peak rotational rate was 0.63×10^{-3} rad/s for the vertical component, much more than that for the horizontal components. The spectra in Figure 4c show that the dominant frequency band in ground acceleration is about 2–5 Hz (horizontal components) while those in Figure 4d show that the dominant frequency band in ground rotation rate is about 2.5–5.5 Hz for the vertical component. Other studies report observations with the R-1 sensor and compare their point measurements of rotation to array-derived area rotations (e.g., Wassermann et al., 2009).

Discussions

Many authors have emphasized the benefits of studying rotational motions (e.g., Twiss et al., 1993; Spudich et al., 1995; Takeo and Ito, 1997; Teisseyre et al., 2006; Trifunac, 2006, 2009; Igel et al., 2007; and Fichtner and Igel, 2009). We discuss some basic issues briefly.

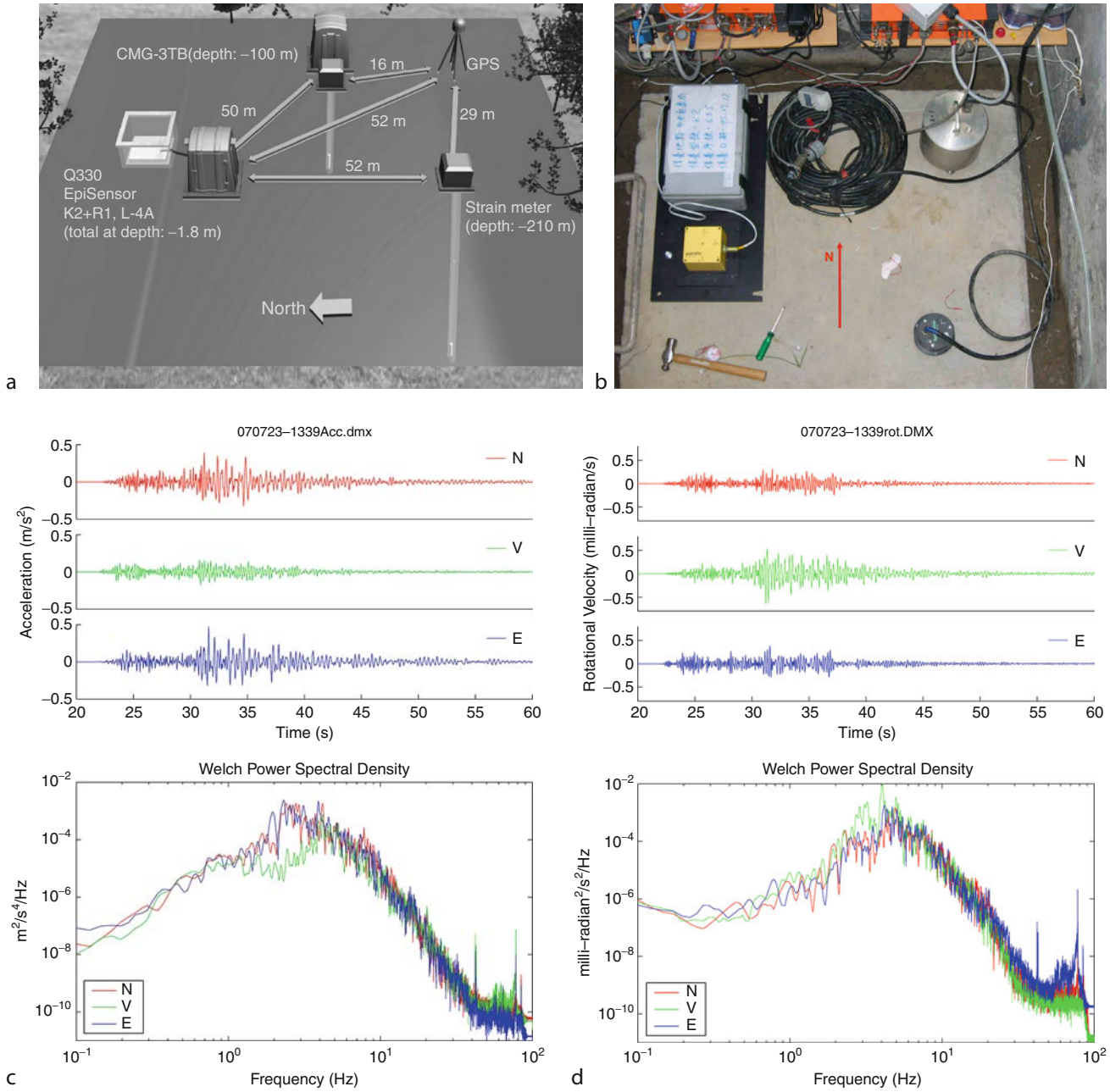
Linear and nonlinear elasticity

Real materials of the Earth are heterogeneous, anisotropic, and nonlinear, especially in the damage zone surrounding faults and in poorly consolidated sediments, soil, and weathered, fractured rock just beneath seismic instruments, particularly typically installed strong-motion instruments. In the presence of significant nonlinearity we are forced to consider the mechanics of chaos (Trifunac, 2009) and to interpret such complexities must record both the rotational and translational components of strong motion.

Seismology is primarily based on the linear elasticity theory, which is applicable to simple homogeneous materials under infinitesimal strain. "Cord" rotation is defined as the curl of the displacement field in Equation 2, and in the classical elasticity theory, the rotational components of motion are contained in the S waves. Meanwhile, continuum mechanics has advanced far beyond the classical theory. In particular, the elasticity theory of the Cosserat brothers (Cosserat and Cosserat, 1909) incorporates (1) a local rotation of continuum particles as well as the translational motion assumed in classical theory, and (2) a couple stress (a torque per unit area) as well as the force stress (force per unit area). In the constitutive equation of classical elasticity theory there are two independent elastic constants while in Cosserat elastic theory there are six or more elastic constants. Pujol (2009) provides a tutorial on rotations in the theories of finite deformation and micropolar (Cosserat) elasticity. Twiss (2009) derives an objective asymmetric micropolar moment tensor from a discrete-block model for a deforming granular material. He also investigates seismogenic deformation associated with volumes of distributed seismicity in three different geographic areas, and finds support in the micropolar model for the effects of a granular substructure on the characteristics of seismic focal mechanisms.

Near-field seismology

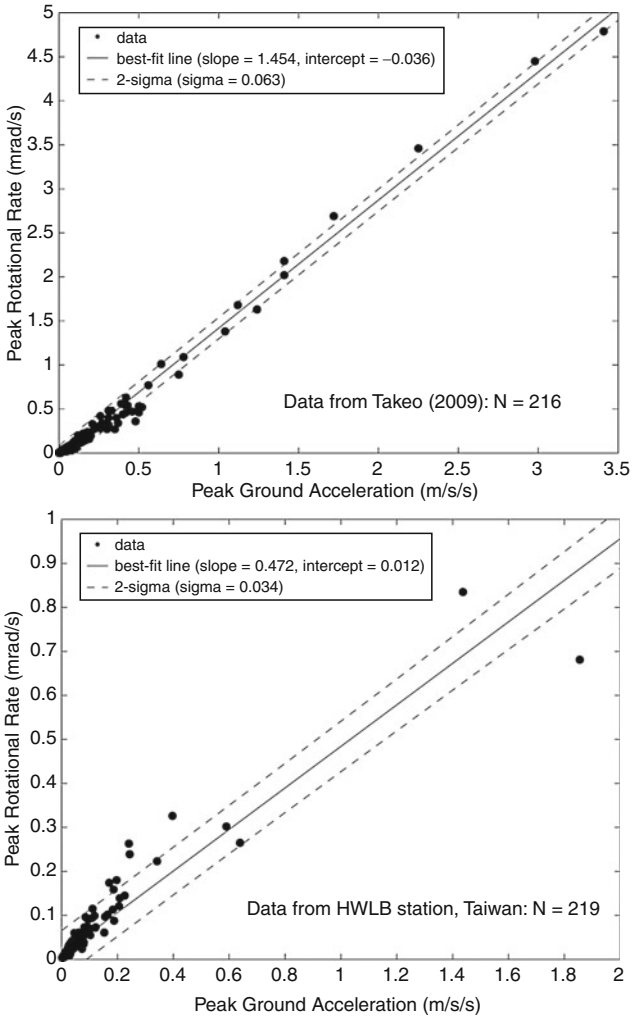
Although the observed rotational motions agree well with the classical elasticity theory for teleseisms, it is not the case for local earthquakes. As first noted by Takeo (1998) and confirmed by Lee et al. (2009b), observed rotational rates from local earthquakes are much larger than those predicted from the classical elasticity theory. For example, Bouchon and Aki (1982) obtained theoretically a maximum rotational rate of 1.5×10^{-3} rad/s for a magnitude 6.5 earthquake, whereas a maximum rotational rate of $>1 \times 10^{-3}$ rad/s for several magnitude 4.5–5.5 earthquakes had been observed in Japan and Taiwan. Takeo (1998) reported the largest rotational rate



Seismology, Rotational, Figure 4 (**a**) Schematic of instrument layout at station HGSD in Taiwan. (**b**) Photo showing the instrument vault containing a K-2 accelerograph, an R-1 rotational seismometer (*yellow*), an EpiSensor accelerometer, and an L-4A velocity seismometer. (**c**) Recorded translational accelerations (*top*) and their spectra (*bottom*) from an M_w 5.1 earthquake 51 km from this site. (**d**) Recorded rotational rates (top) and their spectra (*bottom*) from the same earthquake.

was 2.6×10^{-2} rad/s around the north-south axis from the second largest earthquake (magnitude 5.2) during the 1997 swarm (at 14:09 UTC, 3 March), east of Cape Kawana, offshore Ito, Japan (about 3 km away). As of the end of 2009, the largest rotational rate recorded at the HWLB station (Hualien, Taiwan) was 2.58×10^{-3} rad/s around the east–west axis from a M_w 6.4 earthquake offshore (at UTC 13:02, 19 December 2009) at a hypocentral distance of about 49 km. The peak rotational rate is 1.57×10^{-3} rad/s around the north-south axis, and 0.68×10^{-3} rad/s around the vertical axis. The corresponding peak ground acceleration is 1.16, 1.85, and 0.50 m/s^2 for the east–west, north–south, and vertical component, respectively.

Figure 5 shows the *vertical* peak rotational rate versus *horizontal* peak ground acceleration for the earthquake data set from Takeo (2009) (top frame), and for the earthquake data set (22 August, 2008 to 25 December, 2009) recorded at the HWLB station, Taiwan (bottom frame), where N = number of data points. As noted by Takeo (1998, 2009), there is a reasonable linear relationship in such plots. The data scatter is larger in the Taiwan data than the Takeo (2009) data. This may due to the fact that the earthquake sources for the Takeo (2009) data were from a nearby offshore swarm, whereas the Taiwan data were recorded from many different earthquake sources. Several authors noted similar linear relationships before (see e.g., Spudich and Fletcher, 2008; Stupazzini et al., 2009; Takeo, 2009; and Wang et al., 2009). In particular,



Seismology, Rotational, Figure 5 Vertical peak rotational rate (PRR) versus horizontal peak ground acceleration (PGA) for the earthquake data set from Takeo (2009) (top frame), and for the earthquake data set (22 August, 2008 to 25 December, 2009) recorded at the HWLB station, Taiwan (bottom frame). *N* number of data points.

Takeo (2009) showed "*a linear correlation between the maximum rotational displacements around vertical axis and the maximum* [ground] *velocities*." The two plots in Figure 5 are equivalent to Takeo's linear relationship, without performing the integration for the measured rotational rate and ground acceleration to obtain rotational displacements and ground velocities. The linear slope in Figure 5 has the dimension unit of s/km, or unit for slowness. Spudich and Fletcher (2008) interpreted this "slowness" as the inverse of an "apparent velocity," characterizing the seismic wavefield beneath the recording station.

Spudich and Fletcher (2008) computed peak values of ground strain, torsions, and tilts for the 2004 Parkfield earthquake (M_w 6.0) and four aftershocks (M_w 4.7–5.1) using the data recorded by an array of accelerators. Takeo (2009) noted that his observations of peak rotation values are about 100 times larger for earthquakes with similar magnitudes and distances, and proposed the following explanations: the different spatial scale of rotational motion by a single-point gyro measurement and by an array observation, the effect of topography, and the difference of the degree of maturation between the San Andreas fault and the swarm volume of offshore Ito. Two arrays of both rotational and translational sensors have been deployed by Wu et al. (2009) in Taiwan to study this discrepancy problem, and hopeful a resolution may be found after recording sufficient numbers of earthquakes.

Using explosions to study rotational motions

Since a large earthquake occurring near a station is rare, explosions have been used to study rotational motions by several pioneers (e.g., Graizer, 1991; Nigbor, 1994). Lin et al. (2009) deployed an array of 8 triaxial rotational sensors, 13 triaxial accelerometers, and 12 six-channel, 24-bit dataloggers with GPS time receivers to record two explosions in northeastern Taiwan. These instruments were installed at about 250 m (1 station), 500 m (11 stations), and 600 m (1 station) from the explosions. The 11 stations form a "Center Array" with station spacing of about 5 m. The code name for the first shot with 3,000 kg explosives is "N3P," and that for the second shot with 750 kg explosives is "N3." Although the N3P shot used four times larger amounts of explosives than that used for the N3 shot, the peak ground translational acceleration and rotational velocity at the 13 station sites from the N3P shot are only about 1.5 times larger than that for the N3 shot. Large variations (by tens of %) of translational accelerations and rotational velocities were observed at the very small Center Array. The largest peak rotational rate was observed for the horizontal transverse component: 2.74 $\times$ 10^{-3} and 1.75 $\times$ 10^{-3} rad/s at a distance of 254 m from the N3P and N3 shots, respectively.

The acceleration data from these two explosions were used by Langston et al. (2009) to compute acceleration spatial gradients, horizontal strains and horizontal rotation, and to perform a gradiometric analysis of the strong ground motion wave train. The analysis yields a complex, frequency-dependent view of the nature of seismic wave

propagation over short propagation distances that imply significant lateral velocity changes in the near-surface crustal structure. Areal strain and rotation about the vertical have equal amplitudes and suggest significant wave scattering within the confines of the river valley where the experiment was performed and/or significant departure from an axisymmetric explosion source. Gradiometry shows that the P wave arrives at the array 35° off-azimuth clockwise from the straight-line path and appears to have been refracted from the northern side of the valley. Chi et al. (2011) successfully recovered the first order features of vertical rotational rate ground motions from the translational velocity waveforms in the 0.5–20 Hz bandwidth using the software of Spudich and Fletcher (2009), and strain as large as 10^{-4} was deduced. To fulfill the uniform rotation assumption in the linear elasticity theory, it is necessary to use a small-aperture array. However, inverting data from an array of small spatial dimension requires accurate waveforms of high signal to noise ratio and high sampling rates; since waveforms from adjacent stations are very similar, small noises have strong influence on displacement gradients due to the small station spacing. The Lin et al. (2009) recordings were limited to 200 samples per second, and much higher sampling rate will be necessary to record explosions in the near field.

Processing collocated measurements of translations and rotations

Processing collocated observations of rotation and translation is routinely performed in the inertial navigation units of aircraft and other vehicles. A similar analysis is possible for various combinations of strain components, rotations, and translations. With the exception of velocity–strain combinations (e.g., Gomberg and Agnew, 1996) this terrain was largely unexplored until the work of Lin et al. (2010), who have demonstrated an appropriate set of these equations for earthquake engineering and seismology to recover inertial-frame displacements and rotations. Further, it is already apparent that rotational motions provide useful additional analysis opportunities, simply put, that more data at a site yield more results.

Phase Velocities and propagation directions. A simple calculation for non-dispersive linear-elastic plane waves with transverse polarization shows that the ratio of transverse acceleration to vertical-axis rotation rate is proportional to local phase velocity. This result implies that information on subsurface velocity structure (otherwise only accessible through seismic array measurements and combined analyses) is contained in any single-point measurement that includes rotational sensors. It has been shown that such ratio-derived phase velocities agree with velocities predicted by theory (Igel et al., 2005, Kurrle et al., 2010a). In a recent theoretical study based on full ray theory for Love waves (normal mode summation), Ferreira and Igel (2009) demonstrated that the Love wave dispersion relation also can be obtained by taking the spectral ratio of transverse acceleration to vertical-axis rotation rate. This result implies that seismic shear wave tomography is possible without requiring sub-arrays to determine local mean phase velocities. Information on the direction of propagation also is contained in the azimuth-dependent phase fit between rotations and translations; this fit is optimal in the direction of propagation, from which back azimuths can be estimated to within a few degrees (Igel et al., 2007). Linking observational translations, strains, and rotations together also is advocated by Langston (2007) to yield a snapshot of the wavefield including wave direction, slownesses, and radial/azimuthal amplitude gradients independently at each such station.

Toward a new kind of tomography. The possibility of deriving local dispersion relations from single-station records leads to the question of what subsurface volume one resolves and to what depth velocity perturbations can be recovered. The method of choice to answer this type of question is the adjoint method (Fichtner and Igel, 2009), with which sensitivity kernels (first Fresnel zones) can be calculated to indicate the volume in which the observable (typically travel times) is sensitive to structural perturbations. Fichtner and Igel (2009) introduced a new observable quantity – apparent shear wave velocity – which is a time-windowed ratio of the moduli of translational velocity and rotation angle. It turns out that the sensitivity near the source vanishes, leading to a new type of kernel that shows high sensitivity only in the vicinity of the receiver and in a somewhat smaller portion of that volume than the kernels of translational motions alone. This result implies that a tomographic inversion for near-receiver structures based on rotations and translations is possible and further highlights the potential of rotation measurements. Synthetic tomographic inversions are given in Bernauer et al. (2009).

Scattering properties of the crust: Partitioning of P and S waves. The partitioning of *P* and *S* energy and stabilizing the ratio between the two is an important constraint on the scattering properties of a medium. Igel et al. (2007) discovered surprisingly great rotational energy in a time window prior to teleseismic *S*, containing the *P*-coda. Detailed analysis of the signals and modeling of wave propagation through three-dimensional random media demonstrate that these signals can be explained with *P–SH* scattering in the crust with scatterers of very roughly 5-km correlation length and rms perturbation amplitude of 5% (better constrained). This result further illustrates the efficacy of rotation measurements in their own right, for example, as a filter for SH type motion, as noted by Takeo and Ito (1997).

Similar processing steps will be possible for the horizontal components of rotation and the corresponding components of translation, both to reduce the effects of tilt on the horizontal translational sensors and to extract new information. It is conceivable that the combination of these various components might lead to tight constraints on near-receiver structure, results otherwise only available from array measurements.

Conclusion

Seismology has been very successful in the *far field* because large (magnitude >6) earthquakes occur every week somewhere on Earth, and because classical elasticity

theory works very well for interpreting the recorded *translational* motions at large distances. Because of this success and limited instrumentation options, most funding for earthquake monitoring historically has gone into global and regional seismic networks using only translational seismometers. However, to improve our understandings of damaging earthquakes we must also deploy rotational and translational instruments in the *near field* of active faults where potentially damaging earthquakes (magnitude > 6.5) occur *infrequently*. For strong-motion seismology and engineering, this is a risky business because a damaging earthquake on any given fault may not take place for hundreds of years. Recording ground motions in the near field would require extensive seismic instrumentation along some well-chosen active faults and luck.

Ring laser observations at Wettzell, Germany, and at Piñon Flat, California, demonstrated consistent measurements of rotational ground motions in the *far field*. So far this success can only be demonstrated with one component of rotation. The high cost of present high-precision ring laser gyros makes widespread deployment unlikely. Less expensive and/or less sensitive alternatives are now being pursued by five academic groups (Cowsik et al., 2009; Dunn et al., 2009; Jedlička et al., 2009; Schreiber et al., 2009b; and Takamori et al., 2009). As of the end of 2009, only Taiwan has a modest program (Lee et al., 2009b) to monitor both translational and rotational ground motions from local and regional earthquakes at several permanent seismic stations, as well as by two arrays in a building and a nearby free-field site. These two arrays are designed to "capture" a repeat of the 1906 Meishan earthquake (magnitude 7.1) in the *near field* with both translational and rotational instruments (Wu et al., 2009).

Based on the developments described in the BSSA Special Issue on rotational seismology and engineering (Lee et al., 2009a), observation, analysis, and interpretations of both rotational and translational ground motions will soon play a significant role in seismology and earthquake engineering. An international working group on rotational seismology (IWGoRS) was organized in 2006 to promote investigations of rotational motions and their implications, and for sharing experience, data, software, and results in an open Web-based environment (Todorovska et al., 2008). Anyone can join IWGoRS at http://www.rotational-seismology.org, subscribe to the mailing list, and contribute to the content (publications, data, links, etc.). Rotational seismology is also of interest to physicists using Earth-based observatories for detecting Einstein's gravitational waves (e.g., Lantz et al., 2009) because they must correct for the underlying Earth motion.

Bibliography

Bernauer, M., Fichtner, A., and Igel, H., 2009. Inferring Earth structure from combined measurements of rotational and translational ground motions. *Geophysics*, **74**(6), WCD41–WCD47, doi:10.1190/1.3211110.

Bodin, P., Gomberg, J., Singh, S. K., and Santoyo, M., 1997. Dynamic deformations of shallow sediments in the valley of Mexico, part I: three dimensional strains and rotations recorded on a seismic array. *Bulletin of the Seismological Society of America*, **87**, 528–539.

Bouchon, M., and Aki, K., 1982. Strain, tilt, and rotation associated with strong ground motion in the vicinity of earthquake faults. *Bulletin of the Seismological Society of America*, **72**, 1717–1738.

Chi, W. C., Lee, W. H. K., Lin, C. J., and Liu, C. C., 2011. Inverting horizontal translational for vertical rotational ground motions: a test using data from explosions. [Submitted to the *Bulletin of the Seismological Society of America*, July, 2010.]

Cochard, A., Igel, H., Schuberth, B., Suryanto, W., Velikoseltsev, A., Schreiber, U., Wassermann, J., Scherbaum, F., and Vollmer, D., 2006. Rotational motions in seismology: theory, observation, simulation. In Teisseyre, R., Takeo, M., and Majewski, E. (eds.), *Earthquake Source Asymmetry, Structural Media and Rotation Effects*. Heidelberg: Springer, pp. 391–411.

Cosserat, E., and Cosserat, F., 1909. *Théorie des Corps Déformables*. Paris: Hermann (available from the Cornell University Library Digital Collections) (in French).

Cowsik, R., Madziwa-Nussinov, T., Wagoner, K., Wiens, D., and Wysession, M., 2009. Performance characteristics of a rotational seismometer for near-field and engineering applications. *Bulletin of the Seismological Society of America*, **99**(2B), 1181–1189.

Droste, Z., and Teisseyre, R., 1976. Rotational and displacemental components of ground motion as deduced from data of the azimuth system of seismograph. Publication of the Institute of Geophysics. *Polish Academy of Sciences*, **97**, 157–167.

Dunn, R. W., Mahdi, H. H., and Al-Shukri, H. J., 2009. Design of a relatively inexpensive ring laser seismic detector. *Bulletin of the Seismological Society of America*, **99**(2B), 1437–1442.

Evans, J. R., 2009. Suggested notation conventions for rotational seismology. *Bulletin of the Seismological Society of America*, **99**(2B), 1429–1436.

Farrell, W. E., 1969. A gyroscope seismometer: measurements during the Borrego earthquake. *Bulletin of the Seismological Society of America*, **59**, 1239–1245.

Ferrari, G., 2006. Note on the historical rotation seismographs. In Teisseyre, R., Takeo, M., and Majewski, E. (eds.), *Earthquake Source Asymmetry, Structural Media and Rotation Effects*. Heidelberg: Springer, pp. 367–376.

Ferreira, A., and Igel, H., 2009. Rotational motions of seismic surface waves in a laterally heterogeneous Earth. *Bulletin of the Seismological Society of America*, **99**(2B), 1073–1075.

Fichtner, A., and Igel, H., 2009. Sensitivity densities for rotational ground-motion measurements. *Bulletin of the Seismological Society of America*, **99**(2B), 1302–1314.

Galitzin, B. B., 1912. *Lectures on Seismometry*. St. Petersburg: Russian Academy of Sciences. in Russian.

Gomberg, J., and Agnew, D., 1996. The accuracy of seismic estimates of dynamic strains: an evaluation using strainmeter and seismometer data from Piñon Flat Observatory, California. *Bulletin of the Seismological Society of America*, **86**, 212–220.

Graizer, V. M., 1991. Inertial seismometry methods. Izvestiya of the USSR Academy of Sciences. *Physics of the Solid Earth*, **27**(1), 51–61.

Huang, B. S., 2003. Ground rotational motions of the 1991 Chi-Chi, Taiwan, earthquake as inferred from dense array observations. *Geophysical Research Letters*, **30**(6), 1307–1310.

Igel, H., Schreiber, U., Flaws, A., Schuberth, B., Velikoseltsev, A., and Cochard, A., 2005. Rotational motions induced by the M8.1 Tokachi-oki earthquake, September 25, 2003. *Geophysical Research Letters*, **32**, L08309, doi:10.1029/2004GL022336.

Igel, H., Cochard, A., Wassermann, J., Schreiber, U., Velikoseltsev, A., and Pham, N. D., 2007. Broadband observations of rotational ground motions. *Geophysical Journal International*, **168**, 182–197.

Jedlička, P., Buben, J., and Kozák, J., 2009. Strong-motion fluid rotation seismograph. *Bulletin of the Seismological Society of America*, **99**(2B), 1443–1448.

Kharin, D. A., and Simonov, L. I., 1969. VBPP seismometer for separate registration of translational motion and rotations. *Seismic Instruments*, **5**, 51–66 (in Russian).

Knopoff, L., and Chen, Y. T., 2009. Single-couple component of far-field radiation from dynamical fractures. *Bulletin of the Seismological Society of America*, **99**(2B), 1091–1102.

Kozák, J. T., 2009. Tutorial on earthquake rotational effects: historical examples. *Bulletin of the Seismological Society of America*, **99**(2B), 998–1010.

Kurrle, D., Igel, H., Ferreira, A. M. G., Wassermann, J., and Schreiber, U., 2010a. Can we estimate local Love wave dispersion properties from collocated amplitude measurements of translations and rotations? *Geophysical Research Letters*, **37**, L04307, doi:10.1029/2009GL042215.

Kurrle, D., Igel, H., Ferreira, A. M. G., Wassermann, J., and Schreiber, U., 2010b. First observation of rotational motions from Earth's free oscillations, *Geophysical Research Abstracts*, **12**, EGU2010-3281.

Langston, C. A., 2007. Wave gradiometry in two dimensions. *Bulletin of the Seismological Society of America*, **97**, 401–416.

Langston, C. A., Lee, W. H. K., Lin, C. J., and Liu, C. C., 2009. Seismic-wave strain, rotation, and gradiometry for the 4 March 2008 TAIGER explosions. *Bulletin of the Seismological Society of America*, **99**(2B), 1287–1301.

Lantz, B., Schofield, R., O'Reilly, B., Clark, D. E., and DeBra, D., 2009. Review: requirements for a ground rotation sensor to improve advanced LIGO. *Bulletin of the Seismological Society of America*, **99**(2B), 980–989.

Lee, V. W., and Trifunac, M. D., 1985. Torsional accelerograms. *Soil Dynamics and Earthquake Engineering*, **4**(3), 132–139.

Lee, V. W., and Trifunac, M. D., 1987. Rocking strong earthquake accelerations. *Soil Dynamics and Earthquake Engineering*, **6**(2), 75–89.

Lee, W. H. K., Celebi, M., Igel, H., and Todorovska, M. I., 2009a. Introduction to the special issue on rotational seismology and engineering applications. *Bulletin of the Seismological Society of America*, **99**(2B), 945–957.

Lee, W. H. K., Huang, B. S., Langston, C. A., Lin, C. J., Liu, C. C., Shin, T. C., Teng, T. L., and Wu, C. F., 2009b. Review: progress in rotational ground-motion observations from explosions and local earthquakes in Taiwan. *Bulletin of the Seismological Society of America*, **99**(2B), 958–967.

Lin, C. J., Liu, C. C., and Lee, W. H. K., 2009. Recording rotational and translational ground motions of two TAIGER explosions in northeastern Taiwan on 4 March 2008. *Bulletin of the Seismological Society of America*, **99**(2B), 1237–1250.

Lin, C. J., Huang, H. P., Liu, C. C., and Chiu, H. C., 2010. Application of rotational sensors to correcting rotation-induced effects on accelerometers. *Bulletin of the Seismological Society of America*, **100**, 585–597.

Liu, C. C., Huang, B. S., Lee, W. H. K., and Lin, C. J., 2009. Observing rotational and translational ground motions at the HGSD station in Taiwan from 2004 to 2008. *Bulletin of the Seismological Society of America*, **99**(2B), 1228–1236.

Mallet, R., 1862. *Great Neapolitan Earthquake of 1857*. London: Chapman and Hall, Vol. I and II.

McLeod, D. P., Stedman, G. E., Webb, T. H., and Schreiber, U., 1998. Comparison of standard and ring laser rotational seismograms. *Bulletin of the Seismological Society of America*, **88**, 1495–1503.

Nigbor, R. L., 1994. Six-degree-of-freedom ground motion measurement. *Bulletin of the Seismological Society of America*, **84**, 1665–1669.

Nigbor, R. L., Evans, J. R., and Hutt, C. R., 2009. Laboratory and field testing of commercial rotational seismometers. *Bulletin of the Seismological Society of America*, **99**(2B), 1215–1227.

Oldham, R. D., 1899. Report on the Great Earthquake of 12th June 1897, *Memoir of the Geological Survey of India*, **29**, 379.

Oliveira, C. S., and Bolt, B. A., 1989. Rotational components of surface strong ground motion. *Earthquake Engineering and Structural Dynamics*, **18**, 517–526.

Pham, D. N., Igel, H., Wassermann, J., Cochard, A., and Schreiber, U., 2009. The effects of tilt on interferometric rotation sensors. *Bulletin of the Seismological Society of America*, **99**(2B), 1352–1365.

Pujol, J., 2009. Tutorial on rotations in the theories of finite deformation and micropolar (Cosserat) elasticity. *Bulletin of the Seismological Society of America*, **99**(2B), 1011–1027.

Reid, H. F., 1910. *The Mechanics of the Earthquake*. The California Earthquake of April 18, 1906, Report of the State Earthquake Investigation Commission, Vol. 2. Washington: Carnegie Institution of Washington, pp. 43–47.

Richter, C. F., 1958. *Elementary Seismology.* San Francisco: W. H. Freeman.

Schreiber, K. U., Igel, H., Cochard, A., Velikoseltsev, A., Flaws, A., Schuberth, B., Drewitz, W., and Müller, F., 2005. The GEOsensor project: rotations – a new observable for seismology. In Flury, J., Rummel, R., Reigber, C., and Rothacher, M. (eds.), *Observation of the Earth System from Space*. Heidelberg: Springer, pp. 1–19.

Schreiber, K. U., Hautmann, J. N., Velikoseltsev, A., Wassermann, J., Igel, H., Otero, J., Vernon, F., and Wells, J.-P. R., 2009a. Ring laser measurements of ground rotations for seismology. *Bulletin of the Seismological Society of America*, **99**(2B), 1190–1198.

Schreiber, K. U., Velikoseltsev, A., Carr, A. J., and Franco-Anaya, R., 2009b. The application of fiber optic gyroscopes for the measurement of rotations in structural engineering. *Bulletin of the Seismological Society of America*, **99**(2B), 1207–1214.

Spudich, P., and Fletcher, J. B., 2008. Observation and prediction of dynamic ground strains, tilts, and torsions caused by the M_w 6.0 2004 Parkfield, California, earthquake and aftershocks, derived from UPSAR Array observations. *Bulletin of the Seismological Society of America*, **98**, 1898–1914.

Spudich, P., and Fletcher, J. B., 2009. Software for inference of dynamic ground strains and rotations and their errors from short baseline array observations of ground motions. *Bulletin of the Seismological Society of America*, **99**(2B), 1480–1482.

Spudich, P., Steck, L. K., Hellweg, M., Fletcher, J. B., and Baker, L. M., 1995. Transient stresses at Parkfield, California, produced by the M 7.4 Landers earthquake of June 28, 1992: Observations from the UPSAR dense seismograph array. *Journal of Geophysical Research*, **100**(B1), 675–690.

Stedman, G. E., 1997. Ring laser tests of fundamental physics and geophysics. *Reports on Progress in Physics*, **60**, 615–688.

Stupazzini, M., De La Puente, J., Smerzini, C., Kaser, M., Igel, H., and Castellani, A., 2009. Study of rotational ground motion in the near field region. *Bulletin of the Seismological Society of America*, **99**(2B), 1271–1286.

Suryanto, W., Igel, H., Wassermann, J., Cochard, A., Schuberth, B., Vollmer, D., Scherbaum, F., Schreiber, U., and Velikoseltsev, A., 2006. First comparison of array-derived rotational ground motions with direct ring laser measurements. *Bulletin of the Seismological Society of America*, **96**, 2059–2071.

Takamori, A., Araya, A., Otake, Y., Ishidoshiro, K., and Ando, M., 2009. Research and development status of a new rotational seismometer based on the flux pinning effect of a superconductor. *Bulletin of the Seismological Society of America*, **99**(2B), 1174–1180.

Takeo, M., 1998. Ground rotational motions recorded in near-source region. *Geophysical Research Letters*, **25**(6), 789–792.

Takeo, M., 2009. Rotational motions observed during an earthquake swarm in April, 1998, at offshore Ito, Japan. *Bulletin of the Seismological Society of America*, **99**(2B), 1457–1467.
Takeo, M., and Ito, H. M., 1997. What can be learned from rotational motions excited by earthquakes? *Geophysical Journal International*, **129**, 319–329.
Teisseyre, R., Takeo, M., and Majewski, E. (eds.), 2006. *Earthquake Source Asymmetry, Structural Media and Rotation Effects*. Berlin/Heidelberg: Springer.
Teisseyre, R., Nagahama, H., and Majewski, E. (eds.), 2008. *Physics of Asymmetric Continua: Extreme and Fracture Processes: Earthquake Rotation and Soliton Waves*. Berlin/Heidelberg: Springer.
Todorovska, M. I., and Trifunac, M. D., 1990. Note on excitation of long structures by ground waves. *Journal of Engineering Mechanics-ASCE*, **116**(4), 952–964.
Todorovska, M. I., Igel, H., Trifunac, M. D., and Lee, W. H. K., 2008. Rotational earthquake motions – international working group and its activities. In *Proceedings of the 14th World Conference on Earthquake Engineering*, Beijing, China, October 12–17, Paper ID: S03-02-0031.
Trifunac, M. D., 1979. A note on surface strains associated with incident body waves. *Bulletin of the European Association of Earthquake Engineering*, **5**, 85–95.
Trifunac, M. D., 1982. A note on rotational components of earthquake motions on ground surface for incident body waves. *Soil Dynamics and Earthquake Engineering*, **1**, 11–19.
Trifunac, M. D., 2006. Effects of torsional and rocking excitations on the response of structures. In Teisseyre, R., Takeo, M., and Majewski, E. (eds.), *Earthquake Source Asymmetry, Structural Media and Rotation Effects*. Heidelberg: Springer, pp. 569–582.
Trifunac, M. D., 2009. Earthquake engineering, nonlinear problems in. In Meyers, R. A. (ed.), *Encyclopedia of Complexity and Systems Science*. New York: Springer, pp. 2421–2437.
Twiss, R. J., 2009. An asymmetric micropolar moment tensor derived from a discrete-block model for a rotating granular substructure. *Bulletin of the Seismological Society of America*, **99** (2B), 1103–1131.
Twiss, R., Souter, B., and Unruh, J., 1993. The effect of block rotations on the global seismic moment tensor and patterns of seismic P and T axes. *Journal of Geophysical Research*, **98**, 645–674.
Wang, H., Igel, H., Gallovic, F., and Cochard, A., 2009. Source and basin effects of rotations: comparison with translations. *Bulletin of the Seismological Society of America*, **99**(2B), 1162–1173.
Wassermann, J., Lehndorfer, S., Igel, H., and Schreiber, U., 2009. Performance test of a commercial rotational motions sensor. *Bulletin of the Seismological Society of America*, **99**(2B), 1449–1456.
Wu, C. F., Lee, W. H. K., and Huang, H. C., 2009. Array deployment to observe rotational and translational ground motions along the Meishan fault, Taiwan: a progress report. *Bulletin of the Seismological Society of America*, **99**(2B), 1468–1474.

Cross-references

Body Waves
Earthquake Rupture: Inverse Problem
Earthquake, Magnitude
Earthquakes, Strong-Ground Motion
Great Earthquakes
Propagation of Elastic Waves: Fundamentals
Seismic Hazard
Seismic Instrumentation
Seismological Networks
Surface Waves

SHEAR-WAVE SPLITTING: NEW GEOPHYSICS AND EARTHQUAKE STRESS-FORECASTING

Stuart Crampin
British Geological Survey, Murchison House, Edinburgh, Scotland, UK

Synonyms

Birefringence; Double refraction

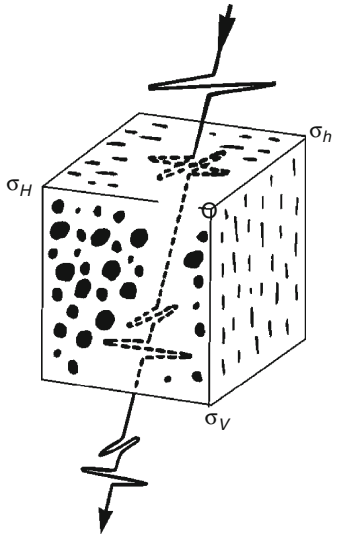
Definition

Observations of shear-wave splitting in the Earth lead to a new understanding of fluid-rock deformation: a New Geophysics, where earthquakes can be stress-forecast.
Shear-wave splitting: Seismic shear-waves propagating through effectively anisotropic solids, such as aligned crystals or aligned microcracks, split into two phases with different velocities and approximately orthogonal polarizations.
New Geophysics: A new understanding of fluid-rock deformation in distributions of stress-aligned fluid-saturated microcracks where microcracks in the crust are so closely spaced that they verge on fracturing and hence are critical-systems (also known as complex-systems). Critical-systems impose fundamentally new properties on conventional subcritical geophysics and hence are a New Geophysics.
Earthquake stress-forecasting: Shear-wave splitting above swarms of small earthquakes can be used as stress-measuring stations to monitor stress-accumulation and stress-relaxation (crack-coalescence) before earthquakes, and hence *stress-forecast* the time, magnitude, and, in some cases, fault-break of impending large earthquakes.

Introduction

Transversely polarized seismic shear-waves propagating through in situ rocks with some form of elastic anisotropy, such as aligned microcracks, split into two orthogonal polarizations which propagate with different velocities, and hence lead to shear-wave splitting. The polarizations are strictly orthogonal for phase-velocity propagation, and approximately orthogonal for group-velocity propagation. Figure 1 is a schematic illustration of shear-wave splitting in the parallel vertical stress-aligned microcracks characteristic of most rocks in the Earth's crust, once below near-surface weathering and stress-release anomalies. Originally referred to as birefringence or double refraction, shear-wave splitting was known to exist from the properties of anisotropic elastic solids (Love, 1927; Nye, 1957).

Double refraction has been observed experimentally as two orthogonally polarized shear-waves with different velocities in a stressed rock sample (Nur and Simmons, 1969). The phenomenon of shear-wave splitting was first



Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting, Figure 1 Schematic illustration of stress-aligned shear-wave splitting through parallel vertical microcracks (after Crampin,1994). σ_V, σ_H, and σ_h are principal axes of vertical stress, and maximum and minimum horizontal stress, respectively.

recognized (and named) in synthetic seismograms in an anisotropic crystalline Earth model (Keith and Crampin, 1977) and in distributions of parallel vertical microcracks (Crampin, 1978). Shear-wave splitting was first positively identified in the field (above swarms of small earthquakes) by Crampin et al. (1980) and in seismic reflection surveys in hydrocarbon reservoirs by Alford (1986). An earlier attempt by Gupta (1973) to demonstrate double refraction in shear-waves some 100 km from small earthquakes failed because of misunderstanding the behavior of shear-wave splitting and the shear-wave window (Crampin et al., 1981).

Although variation of velocity with direction is certainly characteristic of seismic propagation in anisotropic media, shear-wave splitting (or the generalized three-dimensional coupling of surface-wave motion), easily identifiable in particle-motion diagrams (hodograms), is highly diagnostic of some form of effective anisotropy. One of the major difficulties is that observations of shear-wave splitting are path-integration phenomena that give no indication of where the anisotropy is located along the ray path, nor its extent or strength. The only exception is in sedimentary basins where the processing seismic surveys in exploration seismology may locate the anisotropy more accurately (for example, in Angerer et al., 2002).

Note that shear-wave splitting, as in Figure 1, seems quite simple in concept which may tempt authors, such as Gupta (1973), to make unfounded assumptions which lead to incorrect and misleading conclusions. Shear-wave splitting *is* comparatively straightforward, but requires a jump in comprehension from previous isotropic experience. Two early comprehensive reviews (Crampin, 1977, 1981) outlined seismic-wave propagation in anisotropic and cracked elastic media. Crampin and Peacock (2008) recently reviewed the current understanding of shear-wave splitting and identified 17 commonly assumed fallacies in understanding shear-wave splitting that can lead to unjustified conclusions.

It is suggested that the behaviour of shear-wave splitting is now comparatively well-understood. The major new advance is that observations of shear-wave splitting indicate that the stress-aligned fluid-saturated microcracks in almost all rocks are so closely spaced that they verge on failure by fracturing (and earthquakes). Such verging on failure indicates that the distributions of microcracks are critical-systems. Critical systems are a *New Physics* (Davies, 1989), hence a *New Geophysics* that imposes a range of fundamentally new properties on the previous subcritical geophysics. This is a major advance in understanding fluid-rock deformation that has applications to many branches of solid-earth geophysics (Crampin and Peacock, 2005, 2008; Crampin, 2006).

This entry will not discuss the theory of seismic anisotropy and shear-wave splitting which is adequately outlined in Crampin (1977, 1981), Helbig (1994), Maupin and Park (2007), and elsewhere. Here, we merely outline the behaviour of shear-wave splitting in the crust that leads to the New Geophysics and stress-forecasting earthquakes.

Terminology: *references to terminology in the text will be indicated by* (qv).

APE: anisotropic poro-elasticity is a model for deformation/evolution of critical-systems of distributions of stress-aligned fluid-saturated microcracks under changing conditions (Zatsepin and Crampin, 1997; Crampin and Zatsepin, 1997).

Aspect-ratio: Crack thickness over crack diameter.

Band-1 directions: Band-1 ray-path directions are the solid angle between 15° and 45° either side of the crack planes in a distribution of vertical parallel microcracks (Crampin, 1999). APE shows that the effect of increasing (or decreasing) stress is to increase (or decrease) the average time-delays in Band-1 directions.

Band-2 directions: Band-2 ray-path directions are in the solid angle ±15° to the crack planes (Crampin, 1999). Band-2 directions are sensitive to crack density, but APE shows that crack density does not vary consistently for small changes of stress.

Complex-systems: see critical-systems.

Crack density: Dimensionless crack density, *CD,* is specified by $CD = N a^3$ where N is number of cracks of radius a per unit cube.

Critical-systems: Critical-systems (also known as complex-systems, Davies, 1989) are complex heterogeneous interactive systems which verge on fracture-criticality and failure. Extremely common in a huge range of different phenomena, critical-systems impose fundamentally new properties on the previously

subcritical physics and subcritical solid-earth geophysics, and hence are a New Geophysics (Crampin, 2006).

Earthquake prediction: The prediction of time, magnitude, and location of earthquakes by conventional subcritical techniques, but not using shear-wave splitting which implies critical techniques.

Earthquake stress-forecasting: Using shear-wave splitting to monitor stress-accumulation and stress-relaxation in the surrounding rock mass that allows the time and magnitude of earthquakes to be stress-forecast, where other precursory phenomena may indicate the fault-break.

EDA: Extensive-Dilatancy Anisotropy is the name given to the distributions of fluid-saturated microcracks pervading most rocks in the crust (Crampin et al., 1984). EDA-cracks, in ostensibly unfractured rocks, are observed to have crack densities between ∼0.015 and ∼0.045: that is shear-wave velocity anisotropy of ∼1.5 – ∼4.5% (Crampin, 1994). This comparatively small percentage of velocity anisotropy has massive implications for a huge range of phenomena and leads to the New Geophysics.

Fracture-criticality: The value of crack density, ∼0.055, at which stress-aligned microcracks are so closely spaced they verge on fracturing, and hence are critical-systems (qv) (Crampin, 1994, 1999; Crampin and Zatsepin, 1997).

New Geophysics: Shear-wave splitting indicates that microcracks in the crust are so closely spaced they verge on fracture-criticality and fracturing, and hence are critical-systems. Critical-systems are a *New Physics* (Davies, 1989), hence a *New Geophysics*, that imposes a range of fundamentally new properties on conventional subcritical geophysics (Crampin, 2006). Most of these properties have been observed in the field (Crampin and Peacock, 2005, 2008).

PTL-anisotropy: Distributions of horizontal Periodic Thin Layers lead to a form of anisotropy (hexagonal symmetry, commonly known as transverse isotropy, with a vertical axis of symmetry) that is common in sedimentary basins, including many hydrocarbon reservoirs (Wild and Crampin, 1991). Shear-wave velocity anisotropy in PTL-anisotropy may exceed 30% and would be likely to generate cusps in shear-wave velocity sheets. Such cusps have been observed (and modeled) in hydrocarbon reservoirs in the field (Slater et al., 1993).

Seismic anisotropy: Seismic anisotropy describes an elastic solid; these properties vary with direction as in aligned crystals or solids pervaded by aligned microcracks.

Seismic isotropy: Isotropy describes an elastic solid which has the same elastic properties in all directions.

Stress-forecasting earthquakes: see Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting.

TIH- (or HTI-) anisotropy: Transverse isotropy with a horizontal axis of cylindrical symmetry: the typical symmetry of EDA-cracks. Note that although transverse isotropy is specified by five elastic constants, distributions of vertical parallel EDA-cracks are specified by only three: crack density, crack aspect-ratio, and crack strike.

TIV- (or VTI-) anisotropy: Transverse isotropy with a vertical axis of cylindrical symmetry: the typical symmetry of PTL-anisotropy. In TIV-anisotropy, propagation in the horizontal plane is isotropic with no variation with azimuth.

Time-delay: The time-delay between the arrivals of the two split shear-wave phases in seismic anisotropy. Time-delays above earthquakes are frequently normalized by path length to ms/km.

Transverse isotropy: Transverse isotropy (strictly hexagonal symmetry) is the anisotropic symmetry system with five elastic constants, and is isotropic in directions perpendicular to an axis of cylindrical rotational symmetry.

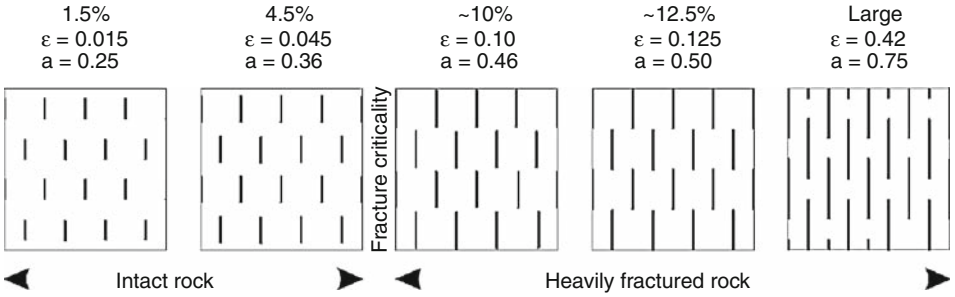
90°-flips in shear-wave polarisations: Abrupt 90° changes in shear-wave polarisations (90°-flips) have been observed and modeled along shear-wave ray paths passing near to point-singularities (see below) (Bush and Crampin, 1991; Crampin, 1991). 90°-flips can also be caused by critically high pore-fluid pressures locally re-orienting stress-aligned fluid-saturated EDA-cracks, both in fluid-injection in critically pressurized hydrocarbon reservoirs (Angerer et al., 2002) and adjacent to seismically-active (critically-pressurized) faults (Crampin et al., 2002).

Fundamental features of shear-wave splitting in the crust

Observations of shear-wave splitting at a horizontal free surface typically display stress-aligned parallel polarisations. Since only transverse isotropy with a horizontal axis of symmetry (TIH-anisotropy) has such parallelism, and only parallel vertical cracks have such symmetry, the splitting is necessarily caused by stress-aligned EDA-microcracks. Microcracks is conformed as temporal changes in time-delays are observed, in fluid-injections (Angerer et al., 2002) and before earthquakes (Crampin and Peacock, 2008). Only microcracks have sufficient compliance to allow such changes.

Observations of azimuthally varying stress-aligned shear-wave splitting, in almost all igneous, metamorphic, and sedimentary rocks in the crust, indicate a minimum of ∼1.5% shear-wave velocity anisotropy and a maximum in ostensibly unfractured rock of ∼4.5% (Crampin, 1994, 1999; Crampin and Peacock, 2008). Since crack density can be written as a dimensionless quantity $CD = N\,a^3$, where N is the number of cracks of radius a per unit cube and if Poisson's ratio is 0.25, $CD \approx 1/100$ of the percentage of shear-wave velocity anisotropy, observed shear-wave splitting can be imaged as uniform distributions of parallel vertical dimensionless cracks in Figure 2 (Crampin, 1994; Crampin and Peacock, 2008).

The evolution of an initially random distribution of stress-aligned fluid-saturated microcracks under changing



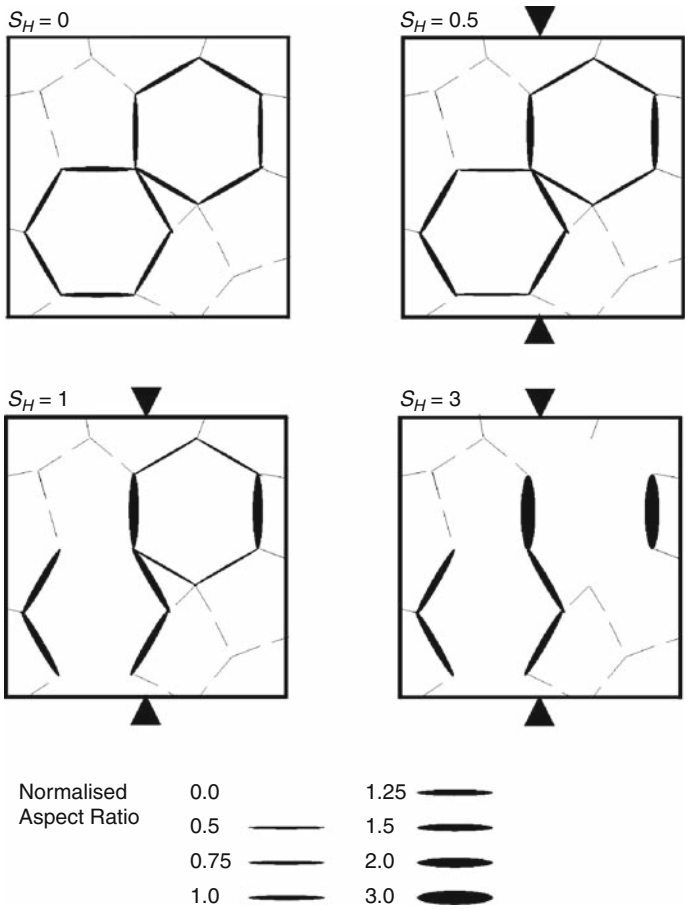
Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting, Figure 2 Schematic illustration of dimensionless three-dimensional crack distributions in in situ rocks for a range of percentages shear-wave velocity anisotropy, where ε is crack density and a is normalized crack radius (after Crampin, 1994). Fracture-criticality is at ~5.5% shear-wave velocity anisotropy (Crampin and Zatsepin, 1997).

conditions can be modeled by the equations of anisotropic poro-elasticity, APE, (Crampin and Zatsepin, 1997). Figure 3 gives a schematic illustration of APE-deformation. The effect of increasing (or decreasing) stress is to increase (or decrease) the average aspect-ratio for cracks parallel to the increasing stress. APE-deformation is almost without parameters and yet matches a huge range of phenomena in exploration and earthquake seismology (Crampin and Peacock, 2008), where the general applicability is because microcracks are so closely spaced that they are critical-systems and lead to the New Geophysics. Note that APE-modeling, because of the universality of critical-systems, is almost without parameters, and yet is widely applicable.

Classes of anisotropic symmetry and shear-wave splitting

Shear-wave splitting has characteristic patterns of polarizations and velocities in three-dimensional propagation depending on the class of anisotropic symmetry (Nye, 1957; Crampin, 1977; Crampin and Kirkwood, 1981). Elastic constants allow eight classes of anisotropic symmetry (including isotropic symmetry which describes a material with two elastic constants, usually written λ and μ, and has identical elastic properties in all directions of propagation). It is the characteristic 3D patterns of planes of mirror symmetry which define the classes of anisotropic symmetry. Four types of anisotropic symmetry are commonly found in the Earth (where the horizontal plane is typically a plane of mirror symmetry):

1. *Orthorhombic symmetry* with nine independent elastic constants has three mutually-orthogonal planes of mirror symmetry. Crystalline olivine, a possible upper mantle constituent has orthorhombic symmetry, where shear-waves split in 3D patterns of polarizations. In addition, simple combinations of EDA- and PTL-anisotropy also have orthorhombic symmetry (Wild and Crampin, 1991). The effects of combinations of EDA- and PTL-anisotropies have been observed in hydrocarbon reservoirs (Bush and Crampin, 1991).
2. *Transverse isotropy with a vertical axis of cylindrical symmetry* commonly referred to as TIV-anisotropy. In TIV, shear-waves split strictly into phases oscillating in the vertical plane (*SV* waves) and phases oscillating horizontally (*SH* waves). There is no azimuthal variation. TIV is the effective anisotropic symmetry of finely divided horizontal layers found in many hydrocarbon reservoirs. It is also the symmetry of small horizontal platelets in shales. Thomsen (1986) derived three highly important approximations (the Thomsen parameters) for the behaviour of seismic waves in TIV media which are directly analogous to parameters measured in reflection surveys. Thomsen parameters are widely used and highly effective in interpreting exploration surveys. Thomsen parameters cannot be used to describe *azimuthally varying* anisotropy, except in isolated directions of sagittal symmetry.
3. *Transverse isotropy with a horizontal axis of cylindrical symmetry* commonly referred to as TIH-anisotropy. TIH is commonly observed above small earthquakes where the faster split shear-wave is polarized parallel to the direction of maximum horizontal stress. Only TIH-anisotropy can produce such parallelism. Only fluid-saturated microcracks are common to the huge variety of geological materials where TIH is observed, and since temporal changes before earthquakes are observed in the field, only fluid-saturated microcracks have sufficient compliance to display temporal changes for small changes in parameters. Consequently, the observed parallelism of fast polarizations directly confirms stress-aligned microcracks throughout almost all rocks in the crust (Crampin and Peacock, 2008).
4. *Monoclinic symmetry* has 13 elastic constants and is the symmetry of two (or more) sets of intersecting non-orthogonal vertical parallel EDA-cracks. Shear-wave splitting in such monoclinic symmetry has been observed and modeled in the field (Liu et al., 1993a). If two intersecting sets of vertical EDA-cracks are orthogonal, the symmetry is orthorhombic. Liu et al. (1993b) display the theoretical effects of intersecting sets of parallel vertical EDA-cracks for a range of parameters and a range of angles of intersection.



Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting, Figure 3 Schematic illustration of anisotropic poro-elastic (APE) modelling of the effects of increasing stress on aspect-ratios of (6% porosity) microcracks, where aspect-ratios are given below (Crampin and Zatsepin, 1995; Crampin, 2006).

Cusps

In strong shear-wave velocity anisotropy, the direction of group-velocity (energy) propagation of the faster of the two split shear-wave may deviate so strongly that they overlap themselves and cause large amplitude cusps. Cusps have been observed and modeled in hydrocarbon reservoirs (Slater et al., 1993). However, the commonest form of cusps that occur in all anisotropic symmetry systems are the cuspoidal lids, fins, and ridges associated with point- and line-singularities in group-velocity shear-wave surfaces (Crampin, 1991; Wild and Crampin, 1991).

Shear-wave window

Shear-waves with *SV* orientations incident on the free surface of an isotropic medium are identical to the incoming shear-wave (but double the amplitude) *only within* the shear-wave window. Outside the window, *S*-wave energy is lost to *SV*-to-*P* conversions so that *SV* waves are heavily distorted and even the arrival of the direct *S* wave may be unreadable. The edge of the shear-wave window is outlined by the critical angle of incidence, $i_c = \sin^{-1}$ (Vs/Vp), where Vp and Vs are the isotropic *P*-wave and *S*-wave velocities, and i_c is the critical angle for *S*-to-*P* conversions (Evans, 1984; Booth and Crampin, 1985). The conical solid angle of the shear-wave window may be considered as *above* small earthquakes or *below* surface recorders.

Other features of the shear-wave window:

1. The critical angle of incidence i_c is $\sim 35^\circ$ for a uniform isotropic medium with a Poisson's ratio of 0.25, where Lamé parameters $\lambda = \mu$.
2. Because of near-surface low-velocity layers, ray paths tend to curve upwards as they approach the free surface, so that the effective shear-wave window is often out to straight-line source-to-receiver incidence of 45° or greater.
3. Topographic irregularities may seriously distort the shear-wave window and may cause *S*-to-*P* conversions (the Local *SP*-waves) both inside and outside the theoretical shear-wave window. Local *SP*-waves are frequently observed in the field as large single-sided pulses on the *SV*-component, and may be mistakenly interpreted as anomalous shear-wave splitting (Crampin, 1990).

4. The shear-wave window only refers to *SV*-propagation. No energy is lost by *SH* waves at any angle of incidence on a plane horizontal free surface. Consequently, only the *SV*-component of any incident-split shear-wave is affected by the shear-wave window.
5. In anisotropic propagation, the shear-wave window may severely disturb shear-wave splitting.

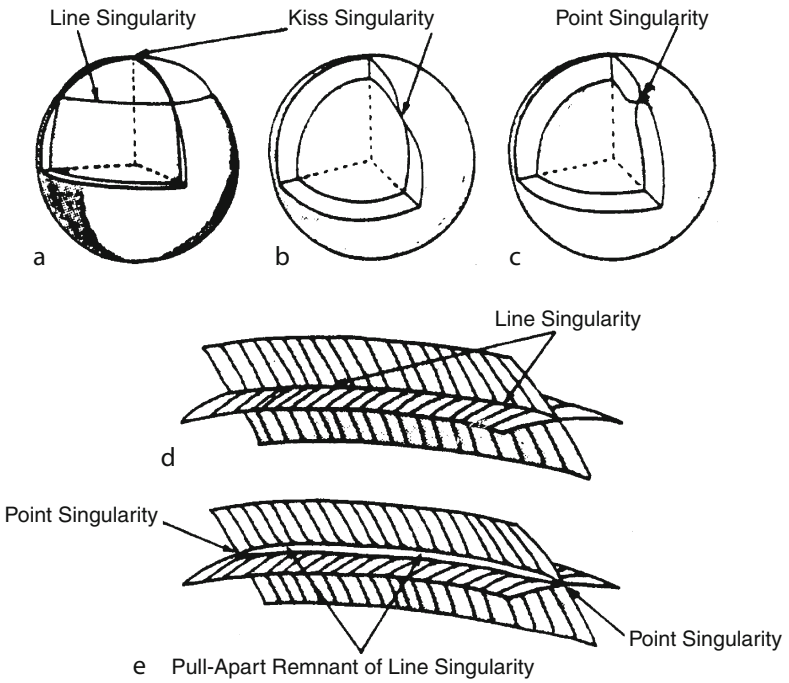
These various anomalies mean that shear-wave splitting recorded at a free surface needs to be interpreted with caution (Crampin and Peacock, 2008).

Shear-wave singularities

The three body waves in anisotropic solids trace out three velocity surfaces for propagation in three-dimensional propagation: a *P* wave; and two orthogonally-polarized shear-wave surfaces. The phase-velocity shear-wave surfaces from a point of propagation necessarily touch and are continuous in a variety of singularities in three-dimensional patterns of direction characteristic of the anisotropic symmetry class (Crampin and Yedlin, 1981; Crampin and Kirkwood, 1981; Wild and Crampin, 1991). The only way for energy of a split shear-wave arrival to pass from one shear-wave surface to the other is along directions passing directly through a singularity.

Figure 4 shows the three types of singularity in *phase-velocity surfaces*:

1. *Kiss-singularities* where two shear-wave surfaces touch tangentially (Figure 4a, 4b). Kiss-singularities occur in directions of principal crystallographic axes and are restricted to cubic, tetragonal anisotropic symmetry, and along the cylindrical symmetry direction of transversely isotropic symmetry. Kiss-singularities can also occur in orthorhombic and monoclinic anisotropic symmetry when two shear-wave surfaces happen to touch serendipitously.
2. *Point-singularities* (sometimes called conical points) where the two shear-wave surfaces touch at the vertices of convex and concave cones (Figure 4c). Shear-waves in all classes of anisotropic symmetry possess point-singularities except transverse isotropy (hexagonal symmetry), which only has kiss- and line-singularities (Figure 4a). Point-singularities are particularly numerous in combinations of PTL- and EDA-anisotropy, where thin layers are pervaded by vertical microcracks (Wild and Crampin, 1991).
3. *Line- or intersection-singularities* occur only in transverse isotropy where the two shear-wave phase-velocity surfaces intersect each other in circular line-(intersection-) singularities about the symmetry axis (Figure 4a, 4d). In even marginally disturbed transverse isotropy, there are dramatic changes. Since both PTL- and EDA-anisotropy have transversely isotropy symmetry with line-singularities, Wild and Crampin



Shear-Wave Splitting: New Geophysics and Earthquake Stress-Forecasting, Figure 4 Schematic illustration of the intersection of two shear-wave phase-velocity surfaces showing topography near: (**a**) line-singularity or intersection-singularity (with kiss-singularity); (**b**) kiss-singularity; and (**c**) point-singularity (after Crampin and Yedlin, 1981). (**d**) A segment of a representation of a line-singularity, and (**e**) a *pull-apart remnant of a line-singularity* when the transverse isotropy symmetry is disturbed (after Wild and Crampin, 1991).

(1991) show that the phase-velocity representations of combinations of PTL and EDA have *pull-apart remnants of line-singularities*. The intersections of the inner and outer shear-wave surfaces separate, and 'pinch' together, in varying degrees of tightness, between a string of point-singularities (Figure 4e).

Singularities in *group-velocity surfaces* have a variety of behaviour:

1. *Kiss-singularities* have simple group-velocity representation, the shear-wave surfaces again touch tangentially.
2. *Point-singularities* in group-velocity surfaces: the faster shear-wave surfaces transform into open almost circular "holes", whereas the slower surfaces transform into flat cuspoidal "lids" which exactly fit into the open holes. The effects of such point-singularities have been observed and modeled in a hydrocarbon reservoir the field (Bush and Crampin, 1991). Crampin (1991) modeled synthetic seismograms for ray path directions passing close to point-singularities. The effects can be dramatic with possibly great amplification or attenuation of amplitudes and abrupt 90° changes in shear-wave polarisations (90°-flips) either side of point-singularities.
3. *Line- or intersection-singularities*. In undisturbed transverse isotropy, the group-velocity representations of line-singularities merely intersect with each sheet showing no evidence of intersection, as in the phase-velocity representation in Figure 4d. However, group-velocity representations of pull-apart remnants of line-singularities in disturbed transverse isotropy typically display thin cuspoidal "fins", "ridges", and "lids" of extraordinary complexity (and beauty) (Wild and Crampin, 1991).

There are two conclusions. The effects of pull-apart remnants of line-singularities are sensitive to very small differences in direction of propagation or microcrack geometry, say. The subtleties of shear-wave propagation near the extensive pull-apart remnants of line-singularities are likely to be the reason for many of the complications frequently observed in field records of shear-waves in sedimentary basins. If correctly interpreted, such features can provide crucial information particularly about the relative proportions of PTL- and EDA-anisotropy (Bush and Crampin, 1991).

Sources of shear-wave splitting

There are three common areas of investigation into shear-wave splitting (and seismic anisotropy).

1. *Shear-wave splitting in the upper mantle*, which is typically assumed to be caused by strain-aligned crystals such as olivine, and is used to demonstrate directions of stress and flow in plate tectonics (reviewed by Savage, 1999).
2. *Shear-wave splitting in hydrocarbon reservoirs* in seismic exploration in industrial seismology, where initially "anisotropy is a nuisance" (Helbig, 1994) because many well-established processing procedures assume isotropic propagation and are no longer valid if the rocks are anisotropic. More recently, however, azimuthally varying shear-wave splitting in TIH-anisotropy is being actively used to determine fracture orientations and improve fluid-flow for hydrocarbon recovery (reviewed by Helbig and Thomsen, 2005).
3. *Shear-wave splitting above small earthquakes* initially stimulated development of the theory (reviewed by Crampin, 1981) and observation (Crampin et al., 1980; Alford, 1986), and is now demonstrating the *New Geophysics* of a crack-critical crust (Crampin, 2006), where time-delays directly monitor low-level fluid-rock deformation at levels of deformation well-below those at which fracturing takes place (Crampin and Zatsepin, 1997).

Remarkably, there has been very little interaction between these three areas of investigation. Each has developed largely independently where papers in one area seldom cite other areas of research. Shear-wave splitting in these three areas of investigation will be discussed separately.

Shear-wave splitting in the upper mantle

Seismic anisotropy in the mantle was first recognized from the incompatibility of Rayleigh and Love surface-wave inversions (Anderson, 1961) and from the velocity anisotropy of horizontally propagating *Pn*-waves in oceanic basins (Hess, 1964; Raitt et al., 1969). Another anisotropic surface-wave phenomenon directly analogous to shear-wave splitting of body waves is the coupling of surface-wave motion of Raleigh and Love modes into Generalized modes varying in three-dimensions (Crampin and King, 1977). Just as shear-wave splitting is highly diagnostic of some form of seismic anisotropy, coupled surface-wave motion is highly diagnostic of anisotropy in the mantle. The pronounced coupling between Second Mode Rayleigh and Second Mode Love waves across Eurasia could be caused by as little as 4 km of 4% shear-wave velocity anisotropy at the top of the upper mantle (Crampin and King, 1977).

Shear-wave splitting of body waves in the mantle only became observable with advances in digital instrumentation. The first reports of such shear-wave splitting were by Ando et al. (1980) who reported time-delays of ~1s on 2s-period shear-waves propagating vertically. This was attributed to 2–4% shear-wave velocity anisotropy in propagation through a magma reservoir, 100–200 km in depth beneath a volcanic area in Japan above the subducting Pacific Plate.

Since that time, shear-wave splitting has been observed extensively in the uppermost 200 km of the mantle and in the D region (Silver and Chan, 1988, 1991), reviewed by Savage (1999), and Long and Silver (2009) and others. These observations are principally of shear-wave splitting in *SKS*-phases, which are frequently large amplitude,

where the *P*-S conversion on exiting from the core provides a known *SV* orientation at the start the final shear-wave path to the surface. Frequencies are typically 0.5–0.25 Hz leading to 2–4% shear-wave velocity anisotropy, similar to the crack-induced shear-wave velocity anisotropy in the crust.

The cause of the shear-wave splitting in the mantle is not wholly resolved. The most frequently cited cause is referred as lattice-preferred orientation (LPO) of the crystallographic axes of elastically anisotropic crystals, such as olivine (Karato and Wu, 1993) or orthopyroxene (Ben-Ismail et al., 2001). These crystals are assumed to be aligned by convective flow where the presence of water, induced by bending of subducting slabs (Faccenda et al., 2009), say, could have major effects on crystal orientations and shear-wave splitting (Karato and Jung, 1998; Katayama et al., 2005).

However, the similarities in the degree of shear-wave splitting and the polarization to splitting in the crust which is certainly caused by EDA-cracks suggests the alternative source of anisotropy of fluid-saturated stress-aligned cracks. There is water in much of the upper mantle. Water lowers the melting point of rock, and when crystalline rock first begins to melt, it melts along grain boundaries. Thus the anisotropy in the mantle could be caused by aligned cracks of films of liquid melt (Crampin, 2003).

Shear-wave splitting in hydrocarbon reservoirs

The effective TIV-anisotropy of finely stratified rocks was demonstrated by Postma (1955) and others, where double refraction (shear-wave splitting) only has *SV*- and *SH*-polarisations. Shear-wave splitting of TIV-anisotropy was not much investigated at that time since the predominant single-component vertical seismometers only recorded *SV* waves.

Azimuthally varying shear-wave splitting in hydrocarbon reservoirs burst onto the seismic exploration industry in a series of some dozen papers in two sessions at the 1986 SEG symposium in Houston. Alford (1986) rotated ("Alford rotations") the polarisations of reflection surveys recorded on two-component horizontal geophones from two-component horizontal vibrators, and showed that seismic shear-wave reflection sections displayed coherent behavior only when both source and receivers were rotated into the preferred (anisotropic) polarizations. These papers stimulated great activity in exploring seismics in order to obtain preferred directions of fluid-flow by locating fracture orientations (reviewed by Helbig and Thomsen, 2005). This interest in anisotropy still continues today but is frequently investigated by wide-angle and other *P*-wave surveys because of the expense of three-component shear-wave surveys including instrumentation as well as the tripling of the quantity of data to be recorded and processed.

Note that azimuthally varying shear-wave splitting in hydrocarbon reservoirs has been typically interpreted by the hydrocarbon industry as being caused by *aligned fractures*, where dimensions of the fractures are unspecified but are expected to be large enough to lead to improved hydrocarbon recovery. In contrast, Crampin and Peacock (2008) suggest that the remarkable compliance of shear-wave splitting to low-level deformation necessarily means that the splitting is caused by stress-aligned fluid-saturated *microcracks*. Since in many circumstances microcracks are parallel to macro-fractures, the difference in interpretation is often immaterial. However, the recent demonstration, both in theory (Chapman, 2003) and observations (Maultzsch et al., 2003), that anisotropy is frequency-dependent and varies with the dimensions of microcracks may change this interpretation.

Shear-wave splitting above small earthquakes

Shear-wave splitting was first positively identified in the crust above a swarm of small earthquakes near the North Anatolian Fault in Turkey in the Turkish Dilatancy Project (TDP1) experiment, (Crampin et al., 1980). This experiment was designed to search for the microcrack-induced shear-wave splitting suggested by Crampin (1978). Similar observations have now been observed above earthquakes worldwide (reviewed by Crampin and Peacock, 2008). However observations above earthquakes need to be recorded by three-component seismic networks within the shear-wave window of the earthquakes. This is a severe restriction as swarms of small earthquakes are scarce and intermittent, and seldom observed by seismic networks. Nevertheless, such swarms of small earthquakes have been used as 'stress-measuring stations' to monitor the effects of stress changes before impending earthquakes (Crampin and Peacock, 2008).

Iceland is an optimum location for studying shear-wave splitting. It is an offset of the Mid-Atlantic Ridge where, uniquely, (two) transform faults run onshore. These transform faults have persistent swarm activity monitored by a state-of-the-art seismic network (Stefánsson et al., 1993) available over the Internet (Volti and Crampin, 2003).

The APE model of microcrack deformation (Crampin and Zatsepin, 1997) shows that increases of stress increase crack aspect-ratios (crack thicknesses), and increases of aspect-ratio can be recognized by increases of average time-delays in Band-1 directions of the shear-wave window (Crampin, 1999), and have been used to stress-forecast earthquakes (Crampin et al., 1999, 2008).

The new geophysics

The shear-wave velocity anisotropy observed in the crust of the Earth indicates that the distributions of stress-aligned fluid-saturated microcracks (EDA-cracks) are so closely spaced they verge on fracturing and hence are critical-systems (Figure 2). Such critical microcracks are the most compliant elements of in situ rock, and indicate profound differences from conventional subcritical solid-earth geophysics. Critical systems are a *New Physics* (Davies, 1989), hence a *New Geophysics*, which imposes

a range of fundamentally new properties on conventional subcritical physics (Crampin and Peacock, 2005; Crampin, 2006). All complex heterogeneous interactive systems are critical-systems, and they are extremely common: the weather, clustering of traffic on roads, the life-cycle of fruit flies, and a huge range of physical phenomena from stellar radiation to quantum mechanics. Since the Earth is an archetypal complex heterogeneous interactive system, it is necessarily expected to be a critical system.

The new geophysical properties include:

1. *Monitorability* – effects of criticality (on the geometry of EDA-cracks) can be monitored within the interior of the crust by shear-wave splitting (Crampin, 1994, 2006).
2. *Calculability* – the evolution of microcrack geometry can be calculated by APE (Crampin and Zatsepin, 1997; Crampin, 1999; Angerer et al., 2002).
3. *Predictability* – if changing conditions are known, the effects on crack geometry can be predicted (as in Angerer et al., 2002).
4. *Controllability* – if the intended effects of some operation (fluid-injection, say) are desired (the opening of specific fractures for fluid-flow, say), the effects can be calculated by APE, and the effects can be controlled by feedback by analyzing shear-wave splitting. Crack aspect-ratios, as in Angerer et al. (2002) where the effects of fluid-injections were calculated, could be controlled by feedback.
5. *Universality* – the changes in critical-systems are pervasive over all available space where appropriate conditions are maintained. The coincidence of observations of 1.5–4.5% shear-wave velocity anisotropy with similar stress-oriented polarizations in all types of rock regardless of porosity, rock types, or geology (Crampin, 1994; Crampin and Peacock, 2005, 2008) is an example of universality in the crack-critical Earth, that is difficult to explain in a conventional subcritical Earth.
6. *Sensitivity* – extreme (butterfly wings) sensitivity to initial conditions. A prime example of such sensitivity is the SMSITES experiment in Iceland where distinctive variations in *P*- and shear-wave travel-time variations were observed at 70 km from swarm activity with an energy equivalent to a $M = 3.5$ earthquake at 70 km distance. A $M = 3.5$ earthquake is small, so these effects are observed at hundreds of times the likely source diameter in a conventional subcritical earth and are a clear demonstration of critical-system sensitivity (Crampin et al., 2003).

Stress-forecasting earthquakes

Shear-wave splitting and the New Geophysics lead to a new understanding of tectonic deformation before earthquakes. Initially, tectonic stress accumulates by interactions at plate boundaries leading to deformation of EDA-cracks which can be monitored by shear-wave splitting. Such stress-accumulations are not initially associated with any particular source-zone and may be observable at great distances from any potential source. As the stress-field approaches, fracture-criticality weaknesses are identified and microcracks begin to coalesce onto the potential slip-plane. There is stress-relaxation, and eventually the impending earthquake occurs (Gao and Crampin, 2004). Logarithms of the durations of increases (stress-accumulation) and decreases (crack-coalescence) are each separately proportional (self-similar) to impending earthquake magnitudes (Crampin et al., 2008), similar to the linearity of the Gutenberg–Richter relationship. Consequently, shear-waves above swarms of small earthquakes can be used as "stress-measuring stations" to monitor such stress-accumulation and stress-relaxation before impending earthquakes.

Peacock et al. (1988) were the first to recognize such changes in shear-wave time-delays (now identified with stress-accumulation), using seismic station KNW as a stress- measuring station, before a $M = 6$ earthquake near the San Andreas Fault in California. A few other examples were found worldwide (reviewed by Crampin and Peacock, 2008), but the significant advances came when persistent seismicity in Iceland become available for analysis online. Volti and Crampin (2003) displayed increases of Band-1 time-delays before five earthquakes in Iceland monitoring stress-accumulation. The time, magnitude, and fault-plane of a fifth earthquake were successfully stress-forecast 3 days before it occurred in a comparatively tight time/magnitude window (Crampin et al., 1999, 2008).

Crack-coalescence (stress-relaxation), before the impending earthquake occurs when the increasing stress-accumulation abruptly begin to decrease, was first recognized by Gao and Crampin (2004). Since then, stress-accumulation has been recognized before 15 earthquakes worldwide, of which nine had sufficient seismicity beneath stress-measuring stations to also display crack-coalescence (Crampin and Peacock, 2008).

Unfortunately, swarms of small earthquakes are far too scarce and irregular to be used as stress-measuring stations for reliable routine stress-forecasting. Reliable stress-forecasting requires three 1–1.5 km-deep borehole Stress-Monitoring Sites (SMSs), where a Downhole-Orbital Vibrator source (DOV) (Leary and Walter, 2005) radiate shear-waves to borehole geophones in Band-1 directions. A prototype SMS in non-optimal source-to-geophone geometry between existing boreholes in Iceland recorded exceptional sensitivity to low-level seismicity equivalent to a $M = 3.5$ earthquake at 70 km-distance (Crampin et al., 2003) demonstrating that SMSs have sufficient sensitivity to stress-forecast times and magnitudes of $M = 5$ earthquakes up to 1,000 km from impending source zones (Crampin and Gao, 2010).

Summary

Shear-wave splitting in the Earth reveals a fundamentally new understanding of fluid-rock deformation, where

APE-modeling shows that modifications to shear-wave splitting directly monitor stress-induced modifications to microcrack geometry. Since if the changing conditions can be specified, the deformation can be calculated and predicted by APE; this is a substantial advance, a New Geophysics, on conventional subcritical solid-earth geophysics. Changes in microcrack geometry can be monitored, calculated, predicted, and in some circumstances potentially controlled. There is extreme sensitivity to initial conditions, and universality, so that effects are widely and uniformly distributed.

The New Geophysics allows stress-accumulation and stress-relaxation (crack-coalescence) before earthquakes to be recognized and impending large earthquakes stress-forecast. There are many important implications and applications. In particular, any solid-earth application that cannot accommodate compliant stress-aligned fluid-saturated EDA-microcracks pervading most in situ rocks is in error and may lead to serious misunderstandings. New Geophysics leading to monitorability; calculability; predictability; potential controllability; universality; extreme sensitivity; and earthquake stress-forecasting, is likely to be the most fundamental advance in solid-earth geoscience for many decades.

Bibliography

Alford, R. M., 1986. Shear data in the presence of azimuthal anisotropy: Dilley, Texas. *56th Annual International Society of Exploration Geophysicists Meeting, Houston, Expanded Abstracts,* 476–379.

Anderson, D. L., 1961. Elastic wave propagation in layered anisotropic media. *Journal of Geophysical Research*, **66**, 2953–2963.

Ando, M., Ishikawa, Y., and Wada, H., 1980. *S*-wave anisotropy in the upper mantle under a volcanic area in Japan. *Nature*, **286**, 43–46.

Angerer, E., Crampin, S., Li, X.-Y., and Davis, T. L., 2002. Processing, modelling, and predicting time-lapse effects of overpressured fluid-injection in a fractured reservoir. *Geophysical Journal International*, **149**, 267–280. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Ben-Ismail, W., Barruol, G., and Mainprice, D., 2001. The Kaapvaal craton seismic anisotropy: petrophysical analyses of upper mantle kimberlite nodules. *Geophysical Research Letters*, **28**, 2497–2500.

Booth, D. C., and Crampin, S., 1985. Shear-wave polarizations on a curved wavefront at an isotropic free-surface. *Geophysical Journal of the Royal Astronomical Society*, **83**, 31–45.

Bush, I., and Crampin, S., 1991. Paris Basin VSPs: case history establishing combinations of matrix- and crack-anisotropy from modelling shear wavefields near point singularities. *Geophysical Journal International*, **107**, 433–447.

Chapman, M., 2003. Frequency dependent anisotropy due to meso-scale fractures in the presence of equant porosity. *Geophysical Prospecting*, **51**, 369–379.

Crampin, S., 1977. A review of the effects of anisotropic layering on the propagation of seismic waves. *Geophysical Journal of the Royal Astronomical Society*, **49**, 9–27.

Crampin, S., 1978. Seismic wave propagation through a cracked solid: polarization as a possible dilatancy diagnostic. *Geophysical Journal of the Royal Astronomical Society*, **53**, 467–496.

Crampin, S., 1981. A review of wave motion in anisotropic and cracked elastic-media. *Wave Motion*, **3**, 343–391.

Crampin, S., 1990. The scattering of shear waves in the crust. *Pure and Applied Geophysics*, **132**, 67–91.

Crampin, S., 1991. Effects of point singularities on shear-wave propagation in sedimentary basins. *Geophysical Journal International*, **107**, 531–543.

Crampin, S., 1994. The fracture criticality of crustal rocks. *Geophysical Journal of the Royal Astronomical Society*, **118**, 428–438. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., 1999. Calculable fluid-rock interactions. *Journal of the Geological Society*, **156**, 501–514. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., 2003. Aligned cracks not LPO as the cause of mantle anisotropy, *EGS-AGU-EUG* joint assembly, nice. *Geophysical Research Abstracts*, **5**, 00205. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., 2006. The new geophysics: a new understanding of fluid-rock deformation. In Van Cotthem, A., et al. (eds.), *Eurock 2006: Multiphysics Coupling and Long Term Behaviour in Rock Mechanics*. London: Taylor and Francis, pp. 539–544. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., and Gao, Y., 2010. A review of earthquake stress-forecasting. *Proceedings of the International Symposium on Earthquake Seismology and Earthquake Prediction, Beijing, 2009, Pure and Applied Geophysics,* in press.

Crampin, S., and King, D. W., 1977. Evidence for anisotropy in the upper mantle beneath Eurasia from generalized higher mode surface waves. *Geophysical Journal of the Royal Astronomical Society*, **49**, 59–85.

Crampin, S., and Kirkwood, S. C., 1981. Velocity variations in systems of anisotropic symmetry. *Journal of Geophysics*, **49**, 35–42.

Crampin, S., and Peacock, S., 2005. A review of shear-wave splitting in the compliant crack-critical anisotropic Earth. *Wave Motion*, **41**, 59–77. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., and Peacock, S., 2008. A review of the current understanding of shear-wave splitting and common fallacies in interpretation. *Wave Motion*, **45**, 675–722. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., and Yedlin, M., 1981. Shear-wave singularities of wave propagation in anisotropic media. *Journal of Geophysics*, **49**, 43–46.

Crampin, S., and Zatsepin, S. V., 1995. Production seismology: the use of shear waves to monitor and model production in a poro-reactive and interactive reservoir. *65th Annual International Society of Exploration Geophysicists Meeting, Houston, Expanded Abstracts,* 199–202.

Crampin, S., and Zatsepin, S. V., 1997. Modelling the compliance of crustal rock: II - response to temporal changes before earthquakes. *Geophysical Journal International*, **129**, 495–506. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Crampin, S., Evans, R., Üçer, B., Doyle, M., Davis, J. P., Yegorkina, G. V., and Miller, A., 1980. Observations of dilatancy-induced polarization anomalies and earthquake prediction. *Nature*, **286**, 874–877.

Crampin, S., Evans, R., Doyle, M., and Davis, J. P., 1981. Comments on papers about shear-wave splitting in dilatancy-induced anisotropy by Gupta, I. N., and by Ryall, A., and Savage, W. U. *Bulletin. Seismological Society of America*, **71**, 375–377.

Crampin, S., Evans, R., and Atkinson, B. K., 1984. Earthquake prediction: a new physical basis. *Geophysical Journal of the Royal Astronomical Society*, **76**, 147–156.

Crampin, S., Volti, T., and Stefánsson, R., 1999. A successfully stress-forecast earthquake. *Geophysical Journal International*, **138**, F1–F5.
Crampin, S., Volti, T., Chastin, S., Gudmundsson, A., and Stefánsson, R., 2002. Indication of high pore-fluid pressures in a seismically-active fault zone. *Geophysical Journal International*, **151**, F1–F5. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.
Crampin, S., Chastin, S., and Gao, Y., 2003. Shear-wave splitting in a critical crust: III - preliminary report of multi-variable measurements in active tectonics. *Journal of Applied Geophysics*, **54**, 265–277. Special Issue. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.
Crampin, S., Gao, Y., and Peacock, S., 2008. Stress-forecasting (not predicting) earthquakes: A paradigm shift? *Geology*, **36**, 427–430. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.
Davies, P., 1989. The new physics: a synthesis. In Davies, P. (ed.), *The New Physics*. Cambridge: Cambridge University Press, pp. 1–6.
Evans, R., 1984. Effects of the free surface on shear wavetrains. *Geophysical Journal of the Royal Astronomical Society*, **76**, 165–172.
Faccenda, M., Gerya, T. V., and Burlini, L., 2009. Deep slab hydration induced by bending-related variations in tectonic pressure. *Nature Geoscience*, **2**, 790–793.
Fouch, M. J., and Rondenary, S., 2006. Seismic anisotropy beneath stable continental interiors. *Physics of the Earth and Planetary Interiors*, **258**, 292–320.
Gao, Y., and Crampin, S., 2004. Observations of stress relaxation before earthquakes. *Geophysical Journal International*, **157**, 578–582. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.
Gupta, I. N., 1973. Premonitory variations in S-wave velocity anisotropy before earthquakes in Nevada. *Science*, **182**, 1129–1132.
Helbig, K., 1994. Foundations of anisotropy for exploration seismics. In *Handbook of Geophysical Exploration, Section 1, Seismic Exploration*. Oxford: Elsevier Science, p. 248.
Helbig, K., and Thomsen, L., 2005. 75-plus years of anisotropy in exploration and reservoir seismics: A historical review of concepts and methods. *Geophysics*, **70**, 9ND–23ND.
Hess, H., 1964. Seismic anisotropy of the uppermost mantle under oceans. *Nature*, **203**, 629–631.
Karato, S.-I., and Jung, H., 1998. Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth and Planetary Science Letters*, **157**, 193–207.
Karato, S.-I., and Wu, P., 1993. Rheology of the upper mantle: a synthesis. *Science*, **260**, 771–778.
Katayama, I., Karata, S.-I., and Brandon, M., 2005. Evidence of high water content in the deep upper mantle inferred from deformation microstructures. *Geology*, **33**, 613–616.
Keith, C. M., and Crampin, S., 1977. Seismic body waves in anisotropic media: synthetic seismograms. *Geophysical Journal of the Royal Astronomical Society*, **49**, 225–243.
Leary, P. C., and Walter, L. A., 2005. Physical model for the downhole orbital vibrator (DOV) – I. Acoustic and borehole seismic radiation. *Geophysical Journal International*, **163**, 647–662.
Liu, E., Crampin, S., Queen, J. H., and Rizer, W. D., 1993a. Velocity and attenuation anisotropy caused by microcracks and macrofractures in multiazimuthal reverse VSPs. *Canadian Journal of Exploration Geophysics*, **29**, 177–188.
Liu, E., Crampin, S., Queen, J. H., and Rizer, W. D., 1993b. Behaviour of shear waves in rocks with two sets of parallel cracks. *Geophysical Journal International*, **113**, 509–517.
Long, M. D., and Silver, P. G., 2009. Shear wave splitting and mantle anisotropy: measurements, interpretations, and new directions. *Surveys in Geophysics*, **30**, 407–461.
Love, A. E. H., 1927. *A Treatise of the Mathematical Theory of Elasticity*. New York: Dover. 1944 reprinted.
Maultzsch, S., Chapman, M., Liu, E., and Li, X.-Y., 2003. Modelling frequency dependent seismic anisotropy in fluid-saturated rock with aligned fractures: implication of fracture size estimation from anisotropic measurements. *Geophysical Prospecting*, **51**, 381–392.
Maupin, V., and Park, J., 2007. Theory and observations - wave propagation in anisotropic media. In Schubert, G. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, pp. 289–321, doi:10.1016/B978-044452748-6/00007-9.
Nur, A., and Simmons, G., 1969. Stress-induced anisotropy in rock: an experimental study. *Journal of Geophysical Research*, **74**, 6667–6674.
Nye, J. F., 1957. *Physical Properties of Crystals*. Oxford: Clarendon.
Peacock, S., Crampin, S., Booth, D. C., and Fletcher, J. B., 1988. Shear-wave splitting in the Anza seismic gap, southern California: temporal variations as possible precursors. *Journal of Geophysical Research*, **93**, 3339–3356.
Postma, G. W., 1955. Wave propagation in a stratified medium. *Geophysics*, **20**, 780–806.
Raitt, R. W., Shor, G. G., Francis, T. J. G., and Morris, G. B., 1969. Anisotropy of the Pacific upper mantle. *Journal of Geophysical Research*, **74**, 3095–3109.
Savage, M. K., 1999. Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting? *Reviews of Geophysics*, **37**, 65–106.
Silver, P. G., and Chan, W. W., 1988. Implications for continental structure and evolution from seismic anisotropy. *Nature*, **335**, 34–39.
Silver, P. G., and Chan, W. W., 1991. Shear wave splitting and subcontinental mantle deformation. *Journal of Geophysical Research*, **96**, 16,429–16,454.
Slater, C., Crampin, S., Brodov, L. Y., and Kuznetsov, V. M., 1993. Observations of anisotropic cusps in transversely isotropic clay. *Canadian Journal of Exploration Geophysics*, **29**, 216–226.
Stefánsson, R., et al., 1993. Earthquake prediction research in the South Iceland Seismic Zone and the SIL Project. *Bulletin Seismological Society of America*, **83**, 696–716.
Thomsen, L., 1986. Weak elastic anisotropy. *Geophysics*, **51**, 1954–1966.
Volti, T., and Crampin, S., 2003. A four-year study of shear-wave splitting in Iceland: 2 Temporal changes before earthquakes and volcanic eruptions. *Geological Society, London, Special Publication*, **212**, 135–149. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.
Wild, P., and Crampin, S., 1991. The range of effects of azimuthal isotropy and EDA-anisotropy in sedimentary basins. *Geophysical Journal International*, **107**, 513–529.
Zatsepin, S. V., and Crampin, S., 1997. Modelling the compliance of crustal rock: I - response of shear-wave splitting to differential stress. *Geophysical Journal International*, **129**, 477–494. Available at file http://www.geos.ed.ac.uk/homes/scrampin/opinion/.

Cross-references

Seismic Anisotropy
Seismic Tomography
Seismic Wave Propagation in Real Media: Numerical Modeling Approaches
Seismic Waves, Scattering
Seismic, Receiver Function Technique
Seismic, Viscoelastic Attenuation

SINGLE AND MULTICHANNEL SEISMICS

Tamás Tóth
Geomega Limited, Budapest, Hungary

Definition

Seismics or Seismic survey. The method of imaging the subsurface with reflected and refracted seismic waves, which are generated by controlled sources and sensed by seismic detectors.

Single-channel seismics. Seismic survey using one detector (channel) for sensing returning seismic waves.

Multichannel seismics. Seismic survey using multiple detectors (channels) at different locations for sensing returning seismic waves generated by each source points.

Introduction

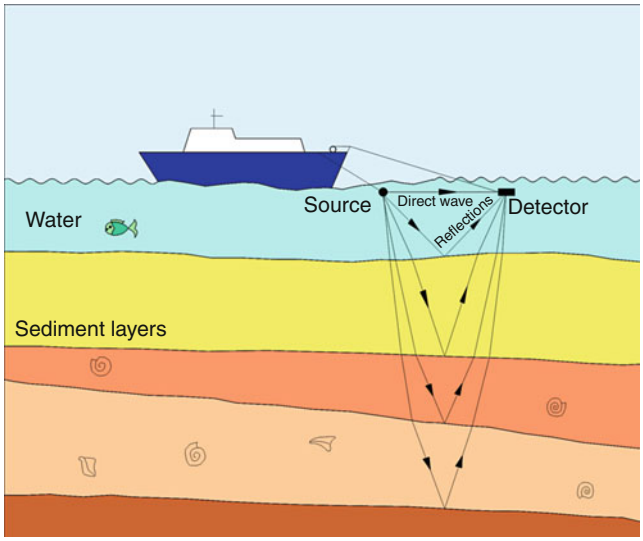
Reflection and refraction seismology has been applied for imaging the subsurface since the 1920s (Karcher, 1987 and Keppner, 1991) and is still one of the most powerful methods of exploration geophysics. Applied technology has evolved considerably over the decades, but the principles have not changed. Imaging is performed by elastic waves generated by controlled sources, such as explosive, vibroseis, air gun, sparker, boomer, or similar sources. Generated waves propagate through the subsurface layers and part of them get reflected or refracted back from layer boundaries. These returning waves are sensed on the surface by seismic detectors, geophones on land, and hydrophones in water.

In case of single-channel seismic recording, elastic waves generated by each shot are recorded by one single detector or a group of detectors. On the other hand, multichannel seismic recording utilizes multiple detectors or groups of detectors located at different positions for recording the wavefield generated by each shot. Multichannel seismics provide several advantages over single-channel seismics; however, simplicity of the recording equipment and high resolution of the recorded seismic profiles make single-channel seismics still a powerful and cost-effective tool for shallow investigations.

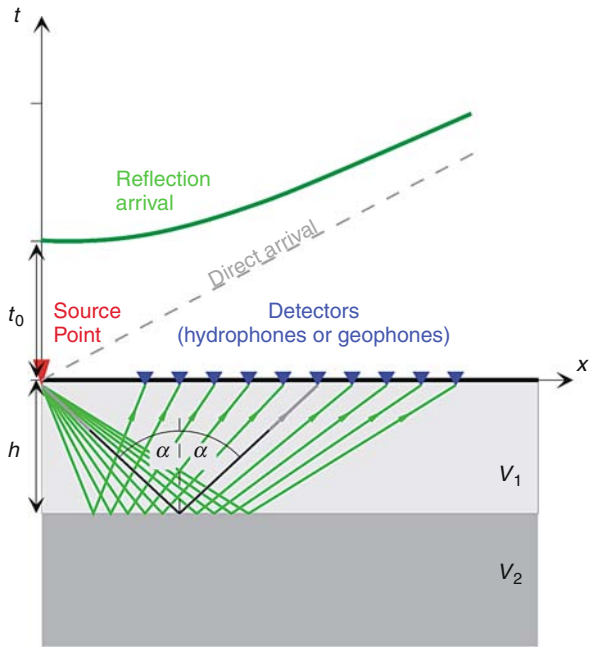
Single-channel seismics

Simplest possible configuration for a seismic reflection survey is single-channel recording. Reflected energy from each shot is sensed by only one receiver or one group of receivers, normally positioned at a constant distance from the source (see Figure 1).

The source–receiver pair is moved, and each shot generates one seismic trace for the profile. In case of single-channel recording, each shot provides information from one subsurface location only, therefore it is critical to have a low cost and high repetition rate source. Shot generation is more labor intensive on land; hence, single-channel recording is hardly used for onshore surveys. Situation is



Single and Multichannel Seismics, Figure 1 During single-channel recording, seismic energy generated by each source point images one subsurface point of each layer boundary, generating one seismic trace.



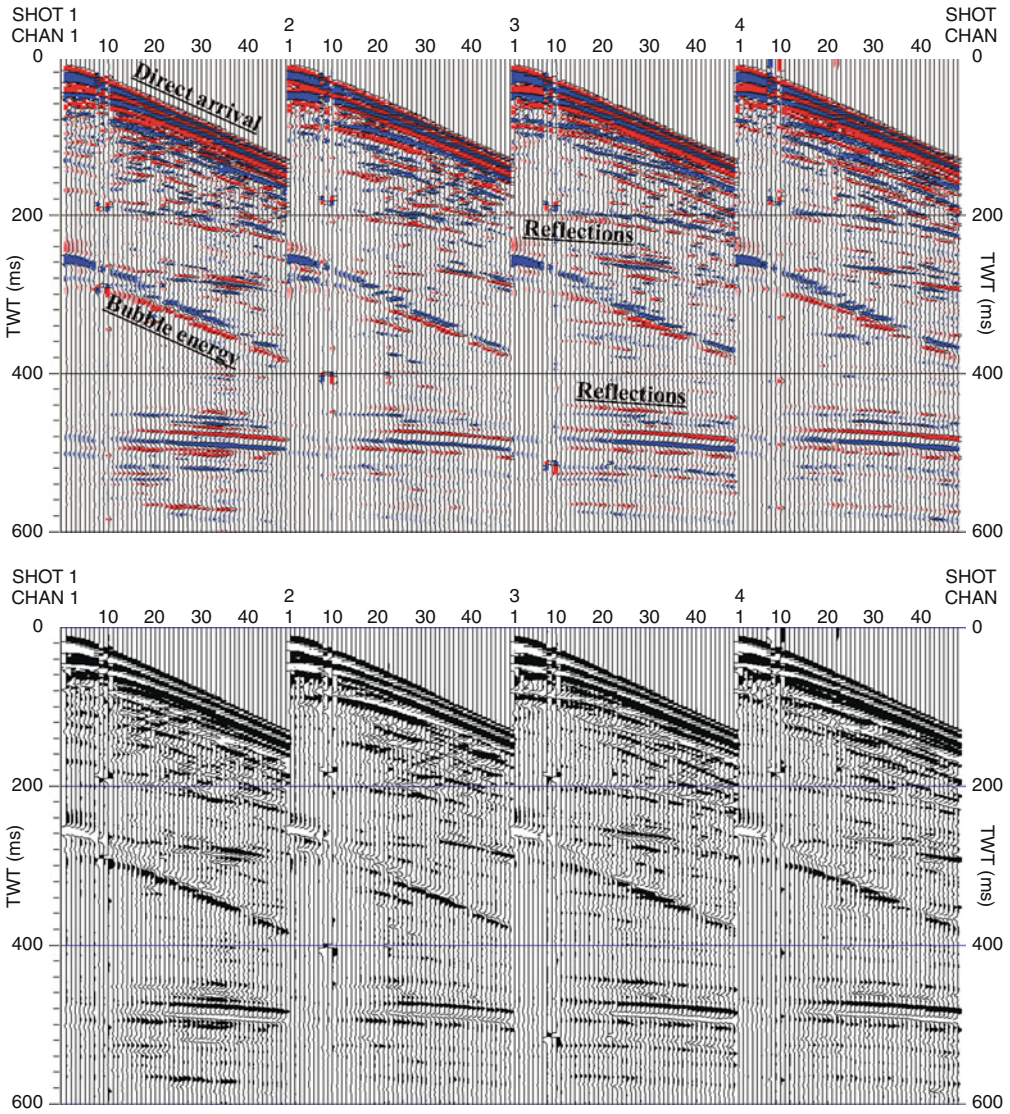
Single and Multichannel Seismics, Figure 2 Reflected energy generated by each source point is recorded by multiple detectors. Each detector images a different subsurface point. Reflection at each subsurface point follows Snell's law. Angles of incidence and reflection (marked α) are equal; hence, distance between the subsurface reflection points is half of the distance between the detectors. Upper part of the Figure shows the theoretical hyperbola of the reflection arrivals and the direct arrival between source point and detectors.

more favorable for offshore surveys, as easy-to-operate sources with high repeatability are available and make single-channel recording an attractive alternative. This is especially true, when relatively shallow "penetration" is sufficient, but *high resolution* is needed.

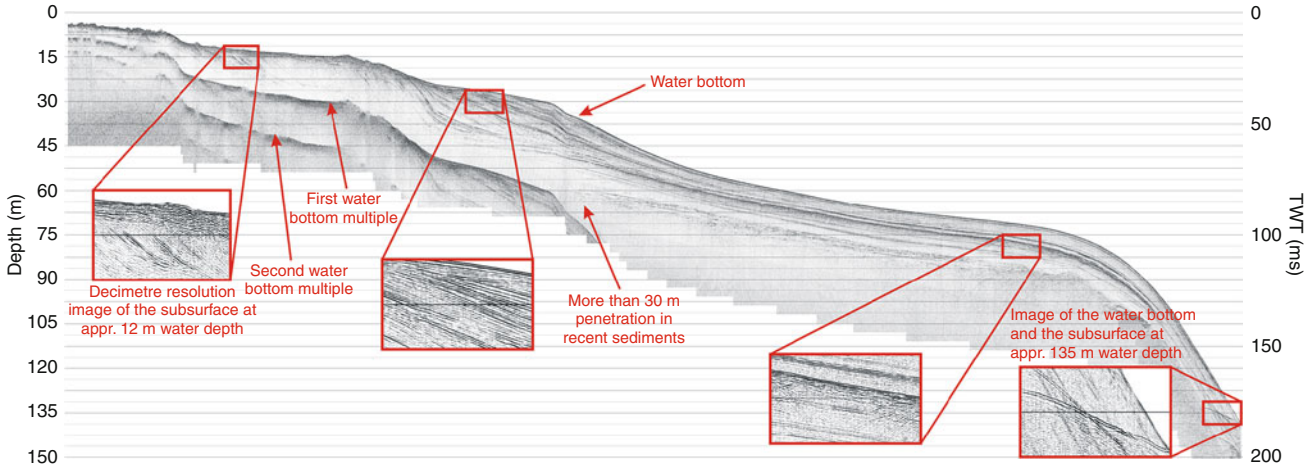
Penetration, on the one hand, strongly depends on the seismic source used; on the other hand, it is also determined by the physical properties of the subsurface strata. Seismic *resolution* is a function of the frequency content of the source used. It is important to note that penetration and resolution are not independent parameters. Higher frequency sources ensure maximum resolution, but as high frequencies attenuate quicker, they provide less penetration. Highest frequency sources are Compressed High Intensity Radar Pulse *(CHIRP)* and boomers, while sparkers and small volume air gun sources with their higher energy provide deeper penetration, but generally lower resolution.

Multichannel seismics

Several advantages are gained by recording the reflected energy of each shot by multiple detectors, which is called multichannel seismic recording (see sketch in Figure 2 and sample shot gathers in Figure 3). The most important of



Single and Multichannel Seismics, Figure 3 The figure shows four subsequent shot gathers recorded by a 48-channel recording system. Upper series displays the shot gathers in red-blue color display with wiggle trace overlay, while the lower series shows the same traces in wiggle trace – varying area (WT/VA) display. Seismic energy was generated by an air gun and the direct arrival, bubble energy and reflections are clearly visible on all of four shot gathers.



Single and Multichannel Seismics, Figure 4 Single-channel boomer profile recorded in the Bay of Napoli showing decimeter resolution and more than 30 m penetration. Comparing the left and right inserts, the attenuation of high frequencies and decreasing of seismic resolution with depth are also visible. High resolution of the seismic combined with gravity coring enabled exceptional dating along the recorded profiles. Details of the research can be found in Sacchi et al., 2009.

these advantages are multifold imaging of the same location, called the common depth point (CDP) and the possibility to estimate seismic velocity from the recorded data itself. Multifold imaging enhances the signal/noise ratio of the final seismic section through stacking. The best possible estimate of velocity field of the subsurface layers is necessary for normal move out *(NMO) correction* and "*migration*" before and after stacking. This information is derived from the parameters of the hyperbola fitted on the reflections in the CDP gathers, a process called "velocity analysis."

Commonly used high-resolution single- and multichannel seismic sources

The CHIRP systems use sweep signals for generating pressure waves. Electromagnetic sweep signals are commonly used in commercial and military radar systems, while acoustic sweeps are used in sonar systems for marine applications, and for *vibroseis* on land. CHIRP systems (e.g., Schock and LeBlanc, 1990; Gutowski et al., 2002) became very popular in marine surveys, while vibroseis is the most commonly used seismic source on land. Marine CHIRP systems typically operate in the 400 Hz–25 kHz frequency range, providing decimeter resolution of the sub-bottom layers down to a few meters or maximum few tens of meters depth in unconsolidated sediments. Advantage of the CHIRP systems is the high repeatability and good S/N (signal-to-noise) ratio. Disadvantage is the loss of phase information of the signal due to the cross-correlation during signal recording. CHIRP sources are used mainly for single-channel seismic recording; however, high-resolution 3D Chirp system has also been developed and used for decimeter-scale object detection (Vardy et al., 2008; Plets et al., 2009).

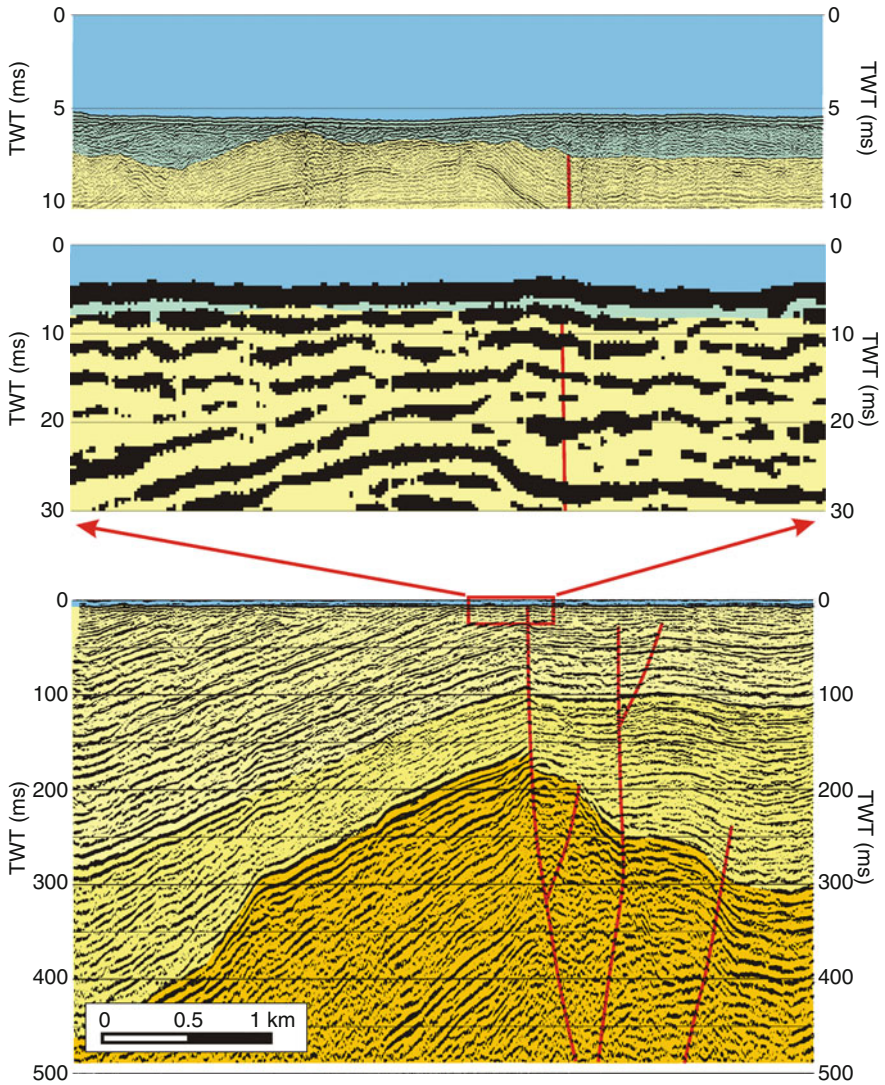
*Boomer*s are electromagnetically driven sources consisting of a flat coil and a metal plate below the coil (Edgerton and Hayward, 1964). Discharge of a high-voltage capacitor bank through the coil generates Eddy-current in the metal plate. The generated Eddy-current violently repels the plate from the coil. Pulling back the plate by strong springs or rubber bands creates a cavitation in the water acting as a sound source. Energy of the source depends on the capacitor bank, which is typically ranging between few 100 and 1,000 J, but several kJ systems have also been used. Boomers generate highly repeatable source signatures in the frequency range of 300 Hz–20 kHz with decimeter resolution and several tens of meters penetration. Innovative system design, such as the IKB-Seistec system utilizing a boomer source and a hydrophone group shielded by a focusing cone (Simpkin and Davis, 1993), can further improve S/N ratio of boomer-based systems providing exceptional combination of resolution and penetration, as shown in Figure 4. Boomers used to be typical single-channel seismic sources; however, multichannel systems developed for high-resolution 3D data acquisition are also built around boomer sources (Müller et al., 2009).

In *sparker* sources, similar to the sparking plug of an engine, discharge of a capacitor bank creates a spark between the positive and negative electrodes of the sparker. This spark vaporizes water between the electrodes and generates a pressure impulse. Energy and shape of the sparker wavelet are influenced by the physical design of the sparker, but is mainly controlled by the capacitance and the voltage of the high-voltage capacitor bank, commonly called "shot box." Typical energy values per shot range between few 100 J and several thousand or even tens of thousands of Joules (Allen, 1972). Large energy sparkers are rarely used nowadays as air gun sources mainly replaced them. Small-energy sparkers are still in use for engineering and research applications as they provide an alternative source fitting between boomers and small air guns. Sparkers

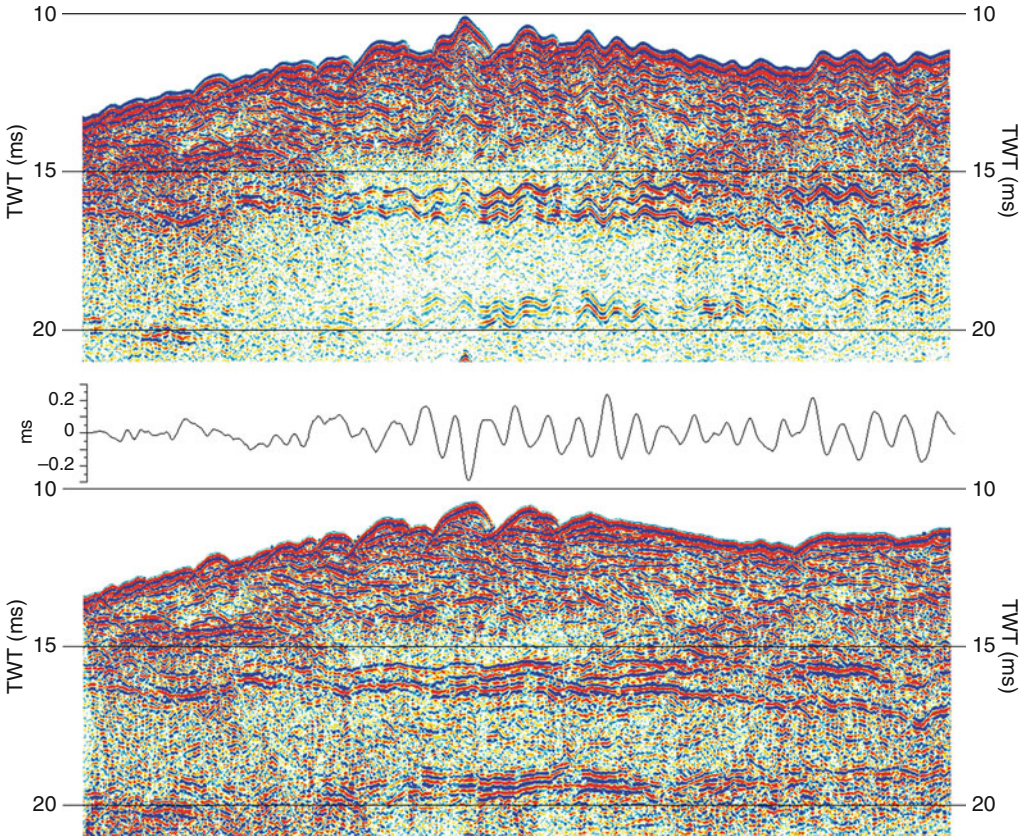
are cost-effective to run, but shape of the wavelet, directivity, and repeatability are better controlled for other sources. Use of the sparker is also limited by the conductivity of the water. It is normally used in saltwater environment as in fresh water, extra care has to be taken to increase the conductivity of the fluid between the electrodes. This can be achieved, for example, by wrapping the sparker in a plastic bag and adding salt inside.

Air gun is a pneumatic seismic energy source generating pressure waves by the sudden release of high-pressure (70–200 bar) air into the water (Giles, 1968). The gun is charged with high-pressure air from a compressor or air container and explosively releases the air from its chamber. The release of the high-pressure air produces a shock wave followed by bubble forming and several oscillations resulting from the expansion and collapse of the air bubble. Good example of bubble energy from the first oscillation is shown in Figure 3.

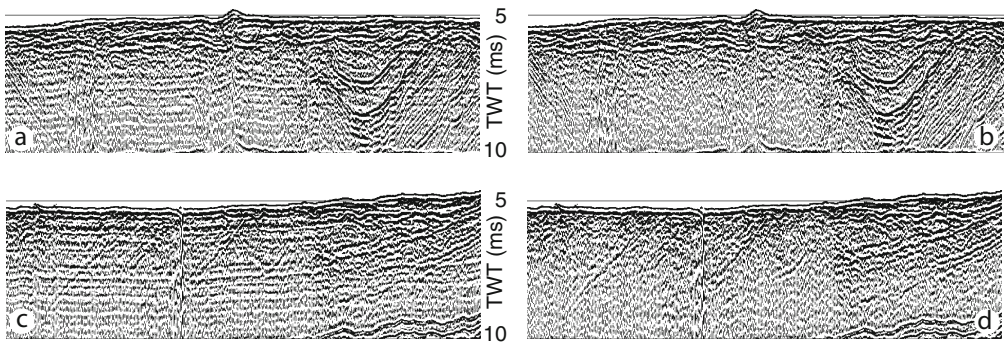
Energy and frequency content of the air gun are determined by the volume of the gun, the pressure of the air and tow depth of the gun. High-resolution surveys require the smallest volume guns, typically 5–20 in^3 (1 in^3 = 16.39 cm^3); however, gun clusters of 1,000–10,000 in^3 are also commonly used during surveys with deeper targets. High-frequency content of small air guns can exceed 1,000 Hz, and large volume guns or gun clusters may



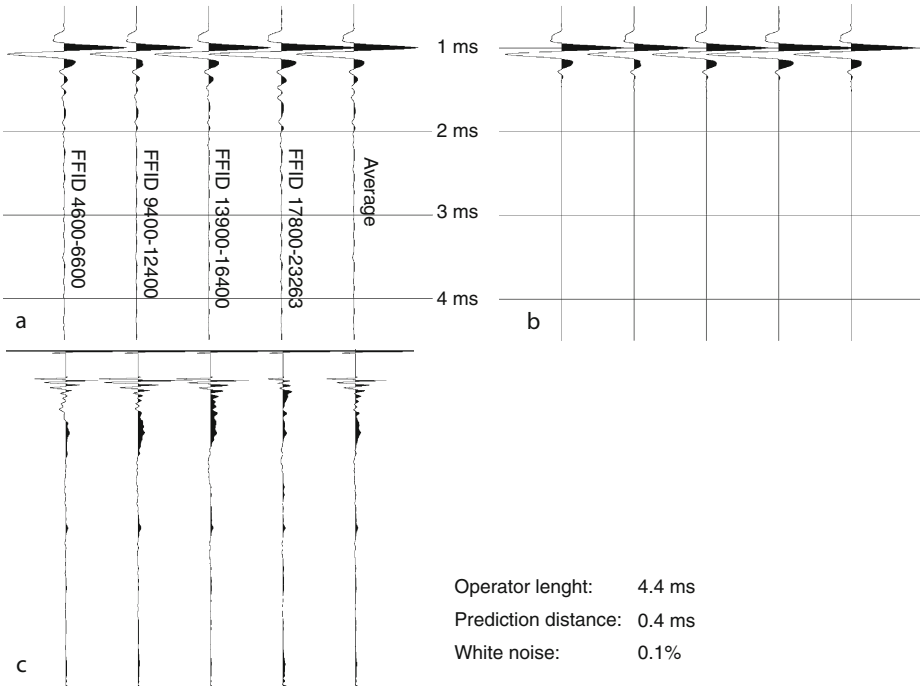
Single and Multichannel Seismics, Figure 5 Bottom image shows a multichannel seismic profile recorded on river Danube, Hungary. Middle image is an optical zoom of the area in the red rectangle. Top image shows the corresponding single-channel boomer profile recorded along the same section of the river. Blue color indicates water, green, recent river sediments, while shades of yellow, Pliocene-Pleistocene sediments. Note the order of magnitude difference in resolution but also in penetration between the two profiles.



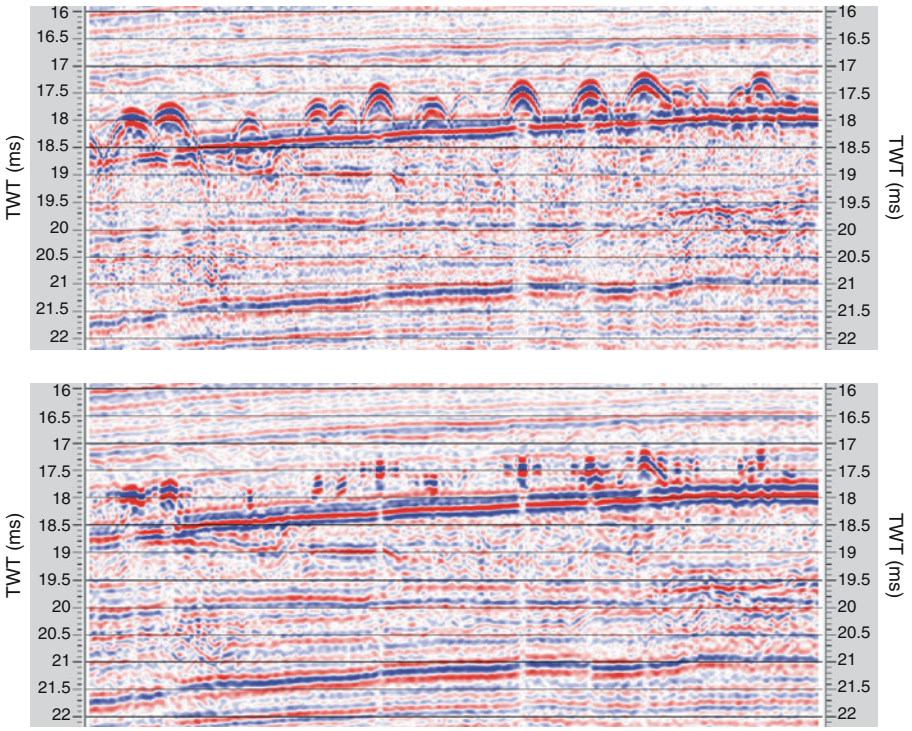
Single and Multichannel Seismics, Figure 6 Top image shows single-channel boomer data recorded on river Rhine with several decimeter amplitude waces. Wave action has been estimated from the cross-correlation of the seismic tarces and the necessary time shift is displayed in the middle graph. Lower image shows the seismic profile after correction for surface wave action. Note that the velocity pull-up and push-down effects due to the velocity difference between water and sediments are clearly visible along the reflector just below 15 ms in the lower image.



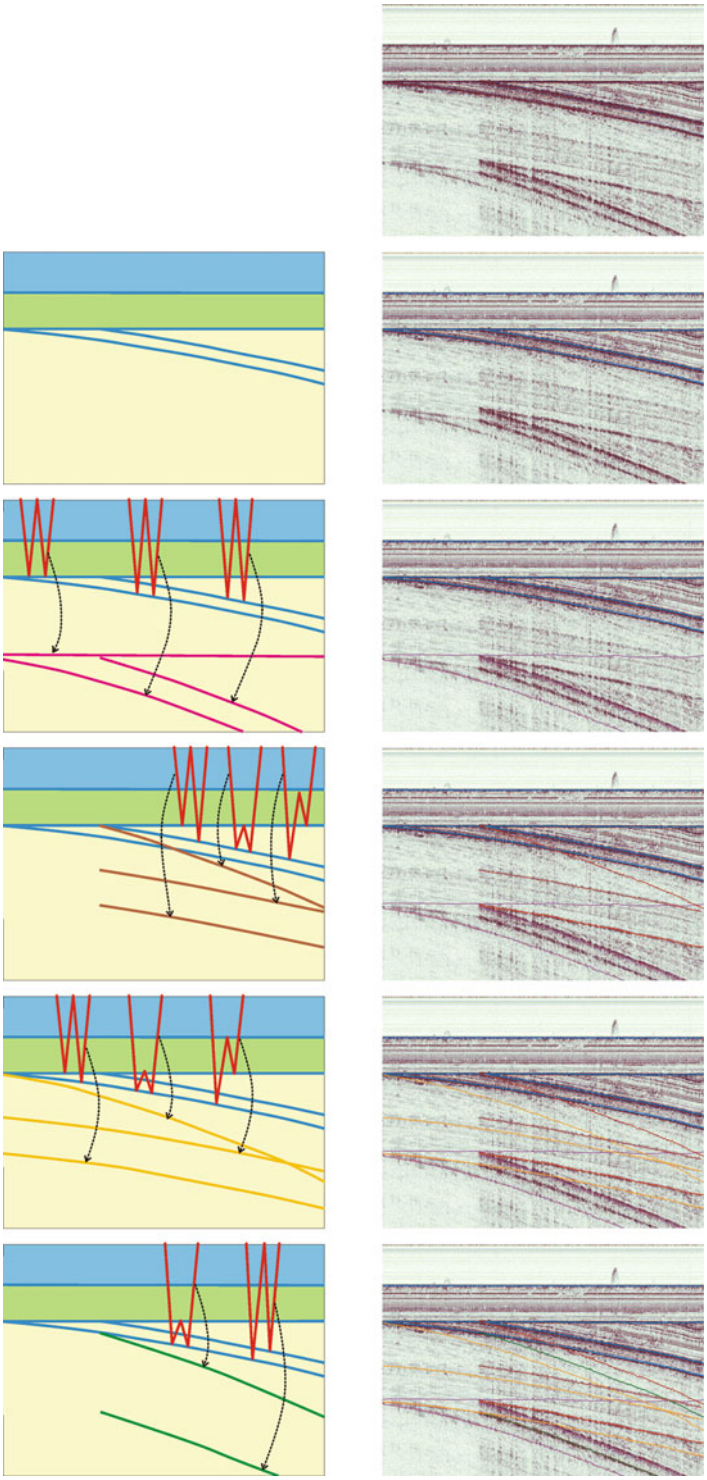
Single and Multichannel Seismics, Figure 7 Two images on the left show profiles without wavelet deconvolution applied. Note the "ringing" of the wavelet running parallel with the water bottom and intersecting genuine reflectors below the water bottom. Profiles on the right have been deconvolved using wavelet-based deterministic deconvolution. The ringing of the wavelet is completely eliminated.



Single and Multichannel Seismics, Figure 8 Wavelets before (**a**) and after (**b**) deterministic deconvolution applied and the deconvolution operator applied (**c**).



Single and Multichannel Seismics, Figure 9 Upper and lower images show the same detail of the seismic profile before and after migration, respectively. Note the collapse of the diffraction hyperbolae around 18 ms TWT. Lateral resolution of the profile is significantly improved by the migration process.



Single and Multichannel Seismics, Figure 10 Top right image shows the original seismic profile recorded in lake Balaton, Hungary. In the lake, approx. 4 m water and 4 m soft, unconsolidated mud can be found above the older Pliocene strata. This, together with the dipping strata in the Pliocene, presents an ideal condition for multiple generation. Top sketch marks with blue the interfaces considered for multiple generation. These are: water surface, water bottom, mud bottom, and two dipping intra-Pliocene strata. First, the free surface multiples (marked by *pink*) are calculated for each of these interfaces. This is followed by different peg-leg multiples calculated and marked by brown, yellow, and green lines on the profile. It is very interesting to observe that very many of the reflections recorded on the seismic profile are actually multiple energies generated by the above mentioned surfaces only.

provide enough energy for several kilometer deep penetration. Air guns are highly repeatable sources with considerable literature on measuring and designing their wavelet (Ziolkowski et al., 1982).

Bubble oscillation can be decreased or even completely diminished by the timely injection of a second volume of high-pressure air. Guns operating under this principle are called *Generated Injection Air Guns* or simply *GI guns*. GI guns have two chambers. The primary chamber (generator) produces the actual pressure impulse, while the secondary chamber (injector) is used to inject a second air volume near the maximum expansion of the first bubble in order to prevent its collapse.

Air gun and *GI* sources due to their higher energy and deeper penetration are mainly used for multichannel recording. A good comparison of single-channel boomer profile and a multichannel air gun profile is shown in Figure 5.

Processing considerations for high-resolution single- and multichannel seismic profiles

Static correction is always a key issue for proper imaging of the subsurface with seismic waves. This is especially true for high-resolution surveys. Offshore seismic surveys present a special static problem due to the wave motion. This equally effects single- and multichannel surveys, but the higher the resolution of the seismics, the more severe the problem can be. Coherency of the reflections in a decimeter resolution boomer survey can be severely affected by surface waves. Heave motion detectors offer a compensation via measuring the wave motions together with the seismic profile; however, if no heave motion is available for the survey, the necessary correction can be calculated from the data. Example of this is shown in Figure 6.

High repeatability of many of the sources (boomer, air guns) offers good opportunity for wavelet determination and application of deterministic *deconvolution* using the calculated wavelet. This is illustrated in Figures 7 and 8 showing two seismic profiles before and after deterministic wavelet deconvolution and the calculated wavelets and deconvolution operators.

Migration is also a key processing step not only for the multichannel, but also for the single-channel seismic profiles. In the later case, estimation of the migration velocity is more complicated, as there is no a priori information available from stacking velocity analysis. Velocity estimates from other measurements or migration velocity analysis can be used for migrating single-channel data. An example of this is shown in Figure 9.

Multiples generated by the free water surface present a significant problem for single-channel seismic profiles. Several processing algorithms exist for multichannel seismic data in order to suppress multiple energy, but almost all of them fail in case of single-channel seismic profiles. Figure 10 shows a single-channel profile recorded in lake Balaton, Hungary with significant multiple energy.

Summary

Seismic surveys provide one of the most detailed image of the subsurface from shallow (few tens of meters) to deep (several kilometers or even tens of kilometers) intervals. Although most of the recent seismic surveys utilize multichannel recording, high-resolution single-channel seismic surveys can also provide a cost-effective solution for offshore surveys. Main advantage of single-channel seismics is high-resolution imaging and relative simplicity of the survey. Multichannel surveys, on the other hand, can significantly improve the signal-to-noise ratio of the seismic profile by applying multifold imaging and at the same time also provide additional information, for example, estimates of the velocity field.

Bibliography

Allen, F. T., 1972. Some characteristics of marine sparker seismic data. *Geophysics*, **37**, 462–470.

Edgerton, H. F., and Hayward, E. C., 1964. The Boomer sonar source for seismic profiling. *Journal of Geophysical Research*, **69**, 3033–3042.

Giles, B. F., 1968. Pneumatic acoustic energy source. *Geophysical Prospecting*, **16**, 21–53.

Gutowski, M., Bull, J., Hestock, T., Dix, J., Hogarth, P., Leighton, T., and White, P., 2002. Chirp sub-bottom profiler source signature design and field testing. *Marine Geophysical Researches*, **23**, 481–492.

Karcher, J. C., 1987. The reflection seismograph: its invention and use in the discovery of oil and gas fields. *The Leading Edge*, **6**(11), 10–19.

Keppner, G., 1991. Ludger mintrop. *The Leading Edge*, **10**(9), 21–28.

Müller, C., Woelz, S., Ersoy, Y., Boyce, J., Jokisch, T., Wendt, G., and Rabbel, W., 2009. Ultra-high-resolution marine 2D-3D seismic investigation of the Liman Tepe/Karantina Island archaeological site (Urla/Turkey). *Journal of Applied Geophysics*, **68**, 124–134.

Plets, R. M. K., Dix, J. K., Adams, J. R., Bull, J. M., Henstock, T. J., Gutowski, M., and Best, A. I., 2009. The use of a high-resolution 3D Chirp sub-bottom profiler for the reconstruction of the shallow water archeological site of the Grace Dieu (1439), River Hamble, UK. *Journal of Archeological Science*, **36**, 408–418.

Sacchi, M., Molisso, F., Violante, C., Esposito, E., Insinga, D., Lubritto, C., Porfido, S., and Tóth, T., 2009. Insights into flood-dominated fan-deltas: very high-resolution seismic examples off the Amalfi cliffed coast, eastern Tyrrhenian Sea. *Geological Society London, Special Publications*, **322**, 33–71.

Schock, S. G., and LeBlanc, L. R., 1990. Chirp sonar: new technology for sub-bottom profiling. *Sea Technology*, **31**(9), 35–43.

Simpkin, P. G., and Davis, A., 1993. For seismic profiling in very shallow water, a novel receiver. *Sea Technology*, **34**(9), 21–28.

Vardy, M. E., Dix, J. K., Henstock, T. J., Bull, J. M., and Gutowski, M., 2008. Decimeter-resolution 3D seismic volume in shallow water: a case study in small-object detection. *Geophysics*, **73**, B33–B40.

Ziolkowski, A., Parkes, G., Hatton, L., and Haugland, T., 1982. The signature of an air gun array. *Geophysics*, **47**, 1413–1421.

Cross-references

Seismic Data Acquisition and Processing
Seismic Velocity-Density Relationships
Seismic Velocity-Temperature Relationships
Seismic, Migration

SLOW EARTHQUAKE

Teruyuki Kato
Earthquake Research Institute, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Definition

An earthquake is a sudden rupture of a fault embedded in the crust or upper mantle, and radiates elastic waves toward the earth's surface to vibrate the ground. Ground vibration includes wide spectrum of frequencies from tens of hertz to hundredth of seconds or lower. The ground vibration is recorded by a pendulum called a seismograph. Since the pendulum has its own frequency characteristics, components of longer periods becomes less visible and are not recognizable in the longest period.

Some types of fault ruptures have predominant component in long-period seismogram and such earthquakes are called slow earthquakes. The earthquake that does not radiate enough energy to be recorded on seismogram is called a silent earthquake or slow slip event (SSE). In this definition, observed pre-seismic or post-seismic slip, or creep events may all be categorized in the silent earthquake or SSE.

Early studies on slow and silent earthquake

Earlier studies of slow earthquake were made by examining difference in magnitude determined by different frequency ranges (e.g., Kanamori, 1972; Kanamori and Stewart, 1979). Seismic magnitude of large earthquakes estimated from shorter period wave such as body waves are significantly smaller than that determined from longer period seismic waves. Ultimately, magnitude determined by seismic wave analysis is often much smaller than the moment magnitude or that estimated by the geodetic data inversion analysis that includes longest period (or permanent displacement) data. These have been interpreted such that the significant amount of moment release was achieved by the component of fault rupture that is slow enough not radiating much seismic wave. Kanamori and Stewart (1979), for example, showed that a series of such slow events were triggered after the June 6, 1960, Chilean earthquake (Mw 9.5).

One of such difference could be manifested by a type of earthquake called "tsunami earthquake." The tsunami earthquake is characterized by anomalously high tsunami generation than that expected from body wave or surface wave magnitude. This can be interpreted by source characteristic such that the fault rupture was slow enough not radiating shorter seismic wave but rapid enough to generate tsunami. The 1896 Sanriku, Japan, earthquake (Ms 7.2, Mw 8.0), 1992 Nicaragua earthquake (Ms 7.0, Mw 7.6) are examples (e.g., Kanamori and Kikuchi, 1993).

If the rupture is much slower and does not radiate any seismic energy, then it is called a silent earthquake. Beroza and Jordan (1990) tested if such very slow or silent earthquake could be detected using spectra of Earth's free oscillations. Another trial is the comparison of long-term moment release rate at subduction zones. If all of strain accumulated along plate boundaries is released by seismic energy, then the released moment at the time of earthquake should be equal to the accumulated moment. Kawasaki et al. (2001) examined this hypothesis and found that the total moment released by earthquakes along the plate boundary is far smaller than the totally accumulated moment in an interseismic stage. This means that there are other unknown slow processes that release remaining moment at the plate boundaries.

One of possible mechanisms of slow strain release would be post-seismic transient displacement. Earlier discovery of post-seismic crustal deformation was found due to the 1946 Nankai, Japan, earthquake (Mw 8.1) (e.g., Okada and Nagata, 1953). The cause of such a post-seismic crustal deformation was interpreted as a retarded slip after the earthquake (Fitch and Scholz, 1971). Kawasaki et al. (1995) asserted that a post-seismic slow rupture was generated after the 1992 Sanriku, Japan, earthquake (Ms 6.9), by examining the strain records observed at the Esashi, Northern Japan, station. However, the question if such post-seismic transients on the records of strain sensors might stem from instrumental drifts after a strong shake was not well resolved.

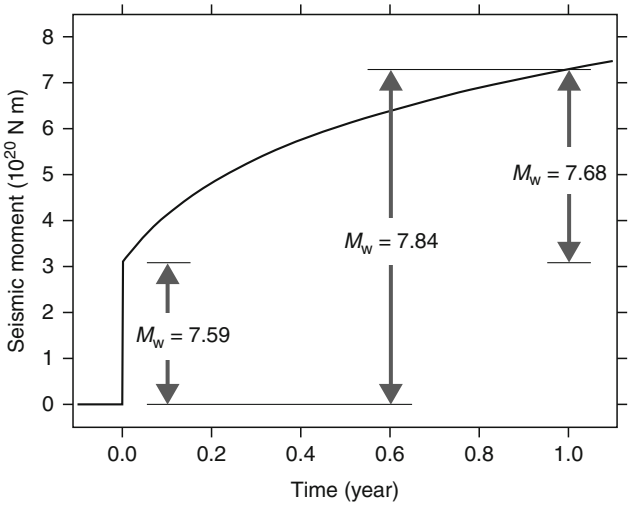
More direct and convincing finding of slow slip events was brought in 1990s by the advent of the Global Positioning System (GPS). Geographical Survey Institute of Japan (GSI) first deployed nationwide GPS network to monitor crustal deformation of the Japanese Islands starting in 1992. The network was much densified in 1996 with more than 600 permanent GPS sites together with near real-time monitoring facility. The network is called GPS Earth Observation Network System (GEONET). It has been further augmented to more than 1,200 sites all over the Japanese Islands. This nationwide array of continuous GPS observation provided such a powerful tool to discover slow slip events around the Japanese Islands, as described in the following section.

Also, the USA deployed a large array of GPS together with strain and tilt meters along the western coast of the North American continent and is called Plate Boundary Observatory (PBO), which also enabled discoveries of slow slip events.

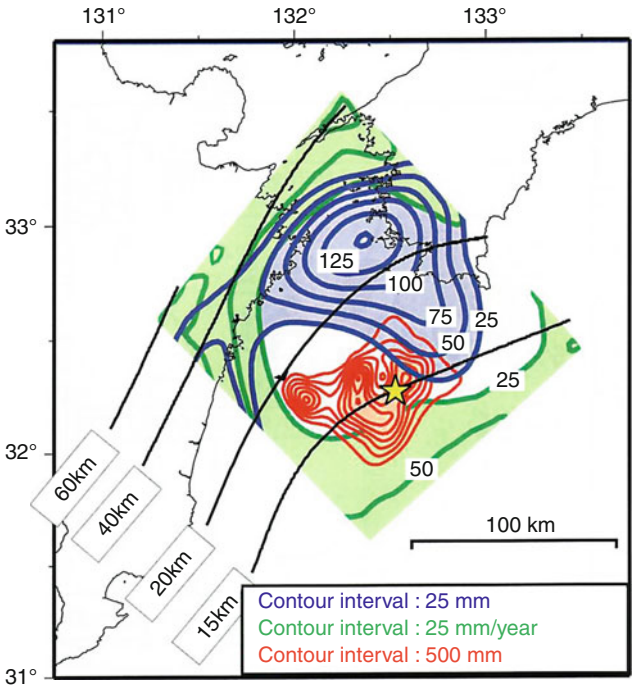
Rapid developments of GPS arrays all over the world, in particular along the Pacific Rim areas, enabled us to discover that silent earthquakes or slow slip events (SSE) are occurring at various areas around the glove, in particular, along the subducting plate interface. Comprehensive description on where do those slow events are taking place are found in Schwartz and Rokosky (2007) and Schwartz (2009).

Long-term slow slip events

A variety of slow slip events, including post-seismic slip, whose time durations are months to years have been discovered along the subducting plate boundaries.



Slow Earthquake, Figure 1 Moment release rate due to the 1994 Sanriku-Haruka-oki, Japan, earthquake. Figure suggests that the moment released in 1 year after the earthquake is bigger than the co-seismic moment release (Heki et al., 1997).



Slow Earthquake, Figure 2 Slip distributions on the subducting plate interface shown by equi-depth contours, projected on the earth's surface: (*Blue*) The slip distribution due to the 1997 slow event at the Bungo Channel, Japan (After Yagi and Kikuchi, 2003). (*Red*) the slip distribution due to the 1968 Hyuganada earthquake, and (*Green*) area of steady slip. (Courtesy by Y. Yagi.)

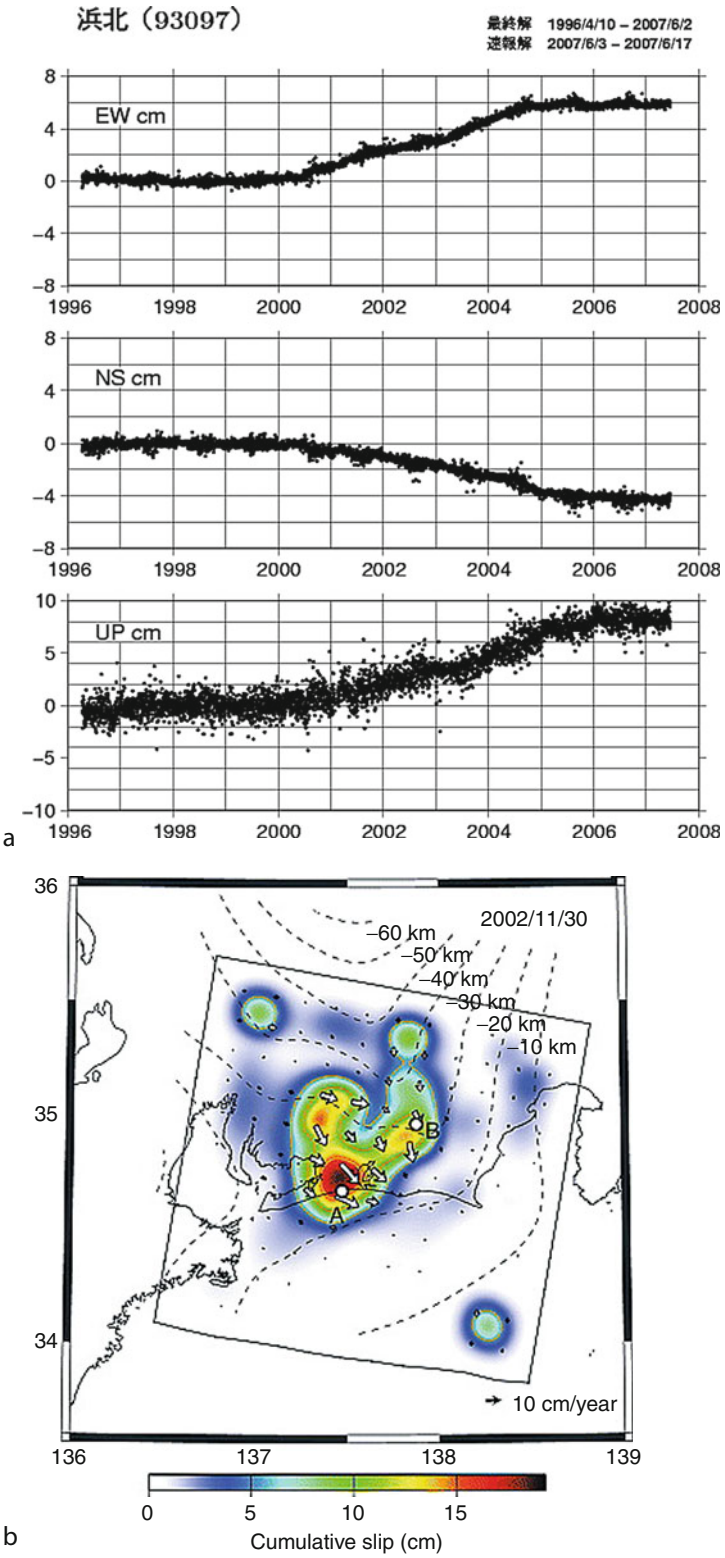
The first eminent case of large slow slip events found using GPS network data would be the post-seismic slip due to the 1994 Sanriku-haruka-oki earthquake (Heki et al., 1997). Heki et al. (1997) analyzed the GEONET data and found that the moment released by the post-seismic slow slips is bigger than that released at the time of earthquake (Figure 1). Then, Hirose et al. (1999) found a slow slip event in the Bungo Channel, western part of Japan, that occurred after the two 1996 Hyuganada earthquakes of Mw 6.6 on October 19 and Mw 6.7 on December 3 that occurred in the south of the region. This slip continued about 1 year in 1997 and was assumed to be due to a slip on the subducting Philippine Sea plate. Their geodetic inversion suggested that the slip occurred on about 60 km × 60 km surface with the maximum fault slip of about 18 cm, which is equivalent to Mw 6.6 earthquake (Figure 2).

Other examples of slow events found in the Japanese region include (1) east off the Boso Peninsula where slow slip events have repeated in 1996 and in 2002 and they were followed by earthquake swarms (Sagiya, 2004, Ozawa et al., 2003) and (2) a post-seismic transient slip after the 2003 Tokachi earthquake (Mw 8.0) (Miyazaki et al., 2004).

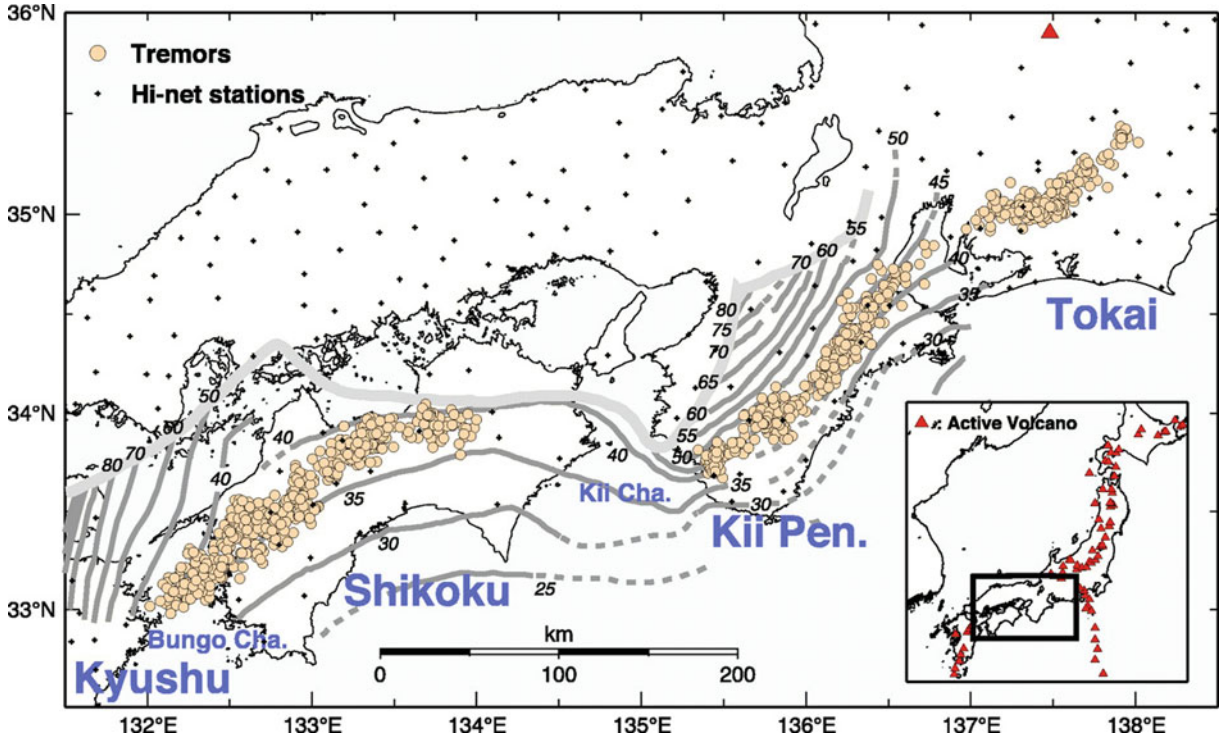
Lowry et al. (2001) found a transient movement at a continuous GPS site established at Cayaco, Guerero region, Mexico in early 1998. The area has been considered as a seismic gap. The event lasted about several months with 2 mm east, 26 mm south, and 16 mm up of displacements. They suggested that the displacement is consistent with a slip along the subduction interface that propagated from east to west. The transient displacement repeated in 2001–2002 for about 6 months. Kostoglodov et al. (2003) and Yoshioka et al. (2004) further investigated the mechanism of the event. Another repeated transient event in Mexico has been found in Oaxaca segment, hundreds of kilometers southeast of Guerrero segment, with more dense GPS array, and was studied by Brudzinski et al. (2007) and Correa-Mora et al. (2008).

Other long-term slow events have been found in Costa Rica (Protti et al., 2004), Alaska-Aleutian (Ohta et al., 2006), and New Zealand (Douglas et al., 2005). SSEs have also been found along the shallower part of the San Andreas fault (e.g., Wesson, 1987; Gladwin et al., 1994; Linde et al., 1996), and in Italy (Crescentini et al., 1999; Amoruso et al., 2002). Readers are asked to refer to Schwartz and Rokosky (2007) and Schwartz (2009) for a detailed review.

Yagi et al. (2001), and Yagi and Kikuchi (2003) used GPS data to examine areal distribution of co-seismic slip, post-seismic slip, and slow slip events for two Hyudanada earthquakes of 1996 (Mw 6.6 and Mw 6.7). They found that these areas complementarily share the areas and postulated that this complementarity may come from different constitutive parameters of slip at respective regions (Figure 2). A similar complementary region of seismic slip and aseismic slip is found also in the northeastern Japan (Yagi et al., 2003). Correa-Mora et al. (2008) indicated that the repeated transients at the Oaxaca segment occurred at the lower extension of the locked part of the



Slow Earthquake, Figure 3 (**a**) A sample GPS time series at the Hamakita site for the Tokai slow slip event. Permanent trend for the period from 1996 to 2000 was removed. Slow slip event was attributed in the period from late 2000 to early 2005. (From webpage of the Geographical Survey Institute) (**b**) Cumulative slip inferred from the extended Network Inversion Filter. *Arrows* show the direction and magnitude of slip of the upper plate (Miyazaki et al., 2006).



Slow Earthquake, Figure 4 Distribution of low-frequency tremor shown in dots (Obara, 2002).

seismogenic zone. These lines of evidence may suggest distinct physical rock properties along the different segments of the subducting plate interface.

The largest and the longest slow slip event, that has ever been found, lasted for about 5 years. It was recorded along the northern margin of the subducting Philippine Sea plate, namely, Tokai district (Figure 3). The event started in the middle of 2000 and continued until mid-2005. Both Ozawa et al. (2002) and Miyazaki et al. (2006) employed a state-space model to delineate temporal evolution of the slow slip event (Segall and Matthews, 1997). The slip amounted to be about several tens of centimeters, which is equivalent to moment magnitude of about 7.2–7.3 earthquake. Figure 3 shows a sample time series of coordinate change at a GEONET station that is located right above the slowly slipping region.

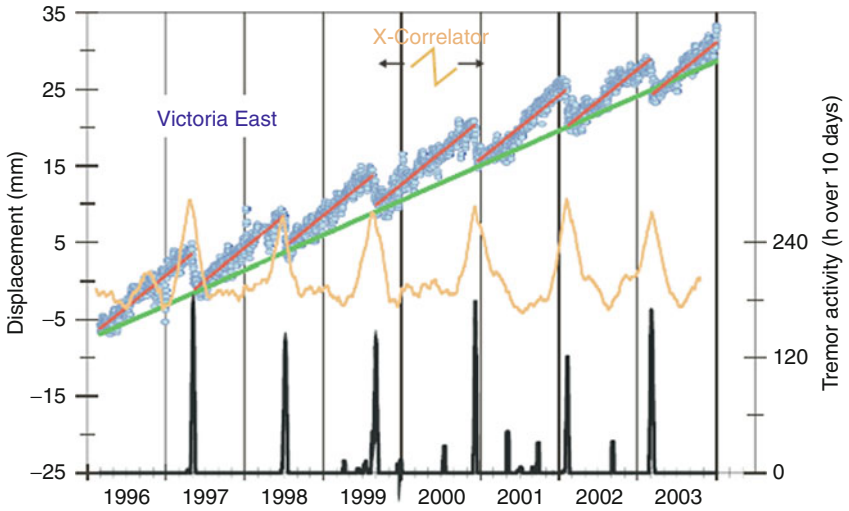
Low-frequency tremor and short-term slow slip

Obara (2002) discovered so-called deep low-frequency tremor using a nationwide seismic array in Japan, which is named as HiNet (High Sensitivity Seismograph Network) constituting of about 600 sites all over Japan. The deep low-frequency tremor is about 1 h of tremor activity, which is located nearly at the plate interface of 35–40 km in depth (Figure 4). The tremor has 0.2–2 Hz in predominant frequency. And then, the tremor was found to be accompanied by small silent slips of a few to several days (Obara and Hirose, 2006; Hirose and Obara, 2006).

Similar nonvolcanic tremor accompanied by short-term slow events were found also in the Northwestern Pacific (Dragart et al., 2001; Rogers and Dragart, 2003; Miller et al., 2002). Dragart et al. (2001) indicated that the SSE rather regularly repeats with interval of about 14 months and is associated with seismic swarm activities (Figure 5). The estimated slip is about 3 cm. Such slow events associated with swarm activities are called episodic tremor and slip (ETS) (Rogers and Dragart, 2003).

Hirose and Obara (2005) found that mid-term (about a month) and short-term (1 or 2 weeks) of slow slip events are accompanied by low-frequency tremors. The short-term silent slip was too small to be monitored by GPS but can be observed on highly sensitive tiltmeters, which are mostly embedded in HiNet boreholes. The estimated amount of slip was only 1–2 cm. They also found repeated similar short-term slow slip events associated with low-frequency tremors at the Tokai region (Hirose and Obara, 2006; Figure 6). Observed slow events were modeled by a slip at the depth of 25 km and the slip amounted only 1.2 cm.

Several lines of evidence from seismological studies suggest that the occurrence of low-frequency tremors and slow slip events may somehow be related to the existence of fluid flow. Shelly et al. (2006) and Obara and Hirose (2006) suggested that episodic tremors that include low-frequency earthquakes are generated by a slow shear slip on the plate interface. Fluids might play a key role for the generation of such low-frequency earthquakes.



Slow Earthquake, Figure 5 Frequently and regularly repeating slow slip events and low-frequency tremors along the Cascadia subduction zones (Dragart et al., 2004).

Matsubara et al. (2009) and Kato et al. (2010) also indicated that the high-pressure fluids released by dehydration from the subducting oceanic crust generates those tremors and slow slip events.

In addition to the low-frequency tremor that occurs at a deeper extension of the seismogenic zone, low-frequency earthquakes (LFE) of a characteristic period of 10–20 s have also been observed at very shallow areas at the updip extension of the seismogenic zones along the subducting plate. They are found along the Nankai trough (e.g., Ishihara, 2003; Obara and Ito, 2005) and at the junction between the Japan trench and the Kuril trench (Asano et al., 2008). Considering that the LFEs along the Nankai trough are occurring within the accretionary prism and have higher dip angle compared with the subucting plate interface, Ito and Obara (2006a) hypothesized that they may occur in the spray faults (or out-of-sequence thrusts) in the wedge.

Existence of fluid would have a key role for generating such low-frequency events, for both deep and shallow low-frequency earthquakes (e.g., Davis et al., 2006; Ito and Obara, 2006b).

Pre-seismic slip

A number of possible pre-seismic slips have been reported. Roeloffs (2006), among others, compared ten distinct examples of reported pre-seismic deformation rate changes before large earthquakes. For example, anomalous crustal deformations were observed before the 1944 Tonankai and 1946 Nankai earthquakes in Japan. Linde and Sacks (2002) showed that these deformations are consistent with the assumed aseismic slip of about 2 m in the down dip extension of the seismically slipped interface. Kanamori and Cipar (1974) and Linde and Silver (1989) postulated that a slow event had preceded the 1960 Chile earthquake.

It might be pointed out that the mechanism of precursory slow slip would be physically the same as slow slip event. Only the difference would be that while the precursory slip is an accelerating phase toward the rapid rupture, slow slip event is the slip that is not grown to a rapid rupture. Studies to clarify the mechanism that creates such difference are yet to be conducted in the future.

Scaling relation

An important question raised is if the discovered slow earthquakes or slow events are just the slow version of "regular" or "normal" earthquakes. If this is the case, those slow events may have the similar scaling relations as the other regular earthquakes. Ide et al. (2007), however, suggested that the slow earthquakes including silent earthquake, slow slip, etc., had different scaling relation and spectral behavior compared with normal earthquakes (Figure 7). They suggested that the characteristic duration (T sec) and the released moment (M_0 Nm) has linear relationship:

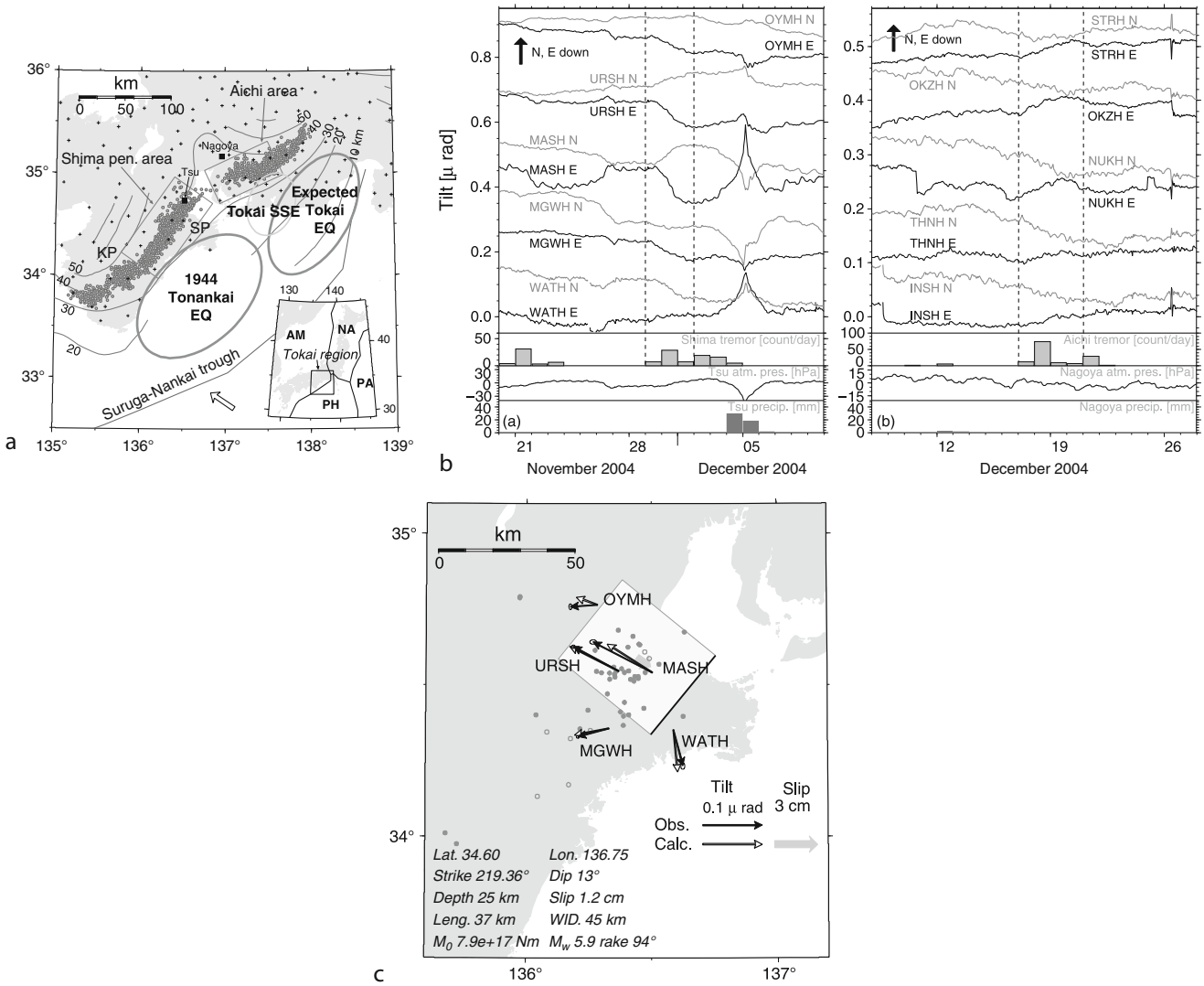
$$M_0 \approx T \times 10^{12-13} \quad (1)$$

While this relation for the regular earthquake is

$$M_0 \approx T^3 \times 10^{15-16} \quad (2)$$

Thus, the characteristics of moment rate spectrum are also different between these events.

This finding may tell us that the slow events are not just a slower extension of seismic slip but rather has a physically different nature. They suggest that such difference would be interpreted by "a constant low-stress drop model or a diffusional constant-slip model" (Ide



Slow Earthquake, Figure 6 (**a**) Map of studied area. (**b**) Observed tilt changes in November and December 2004, due to short-term slow slip events in the Tokai region. Periods of slow events are bounded by *dashed lines*. Recorded low-frequency tremors are also shown. (**c**) Modeled source region and estimated parameters. Observed and predicted tilt changes are shown in vectors (Hirose and Obara, 2006).

et al., 2007). Though the idea is yet to be tested, this might bring us a new insight for understanding the mechanism of fault slip or "earthquake" itself.

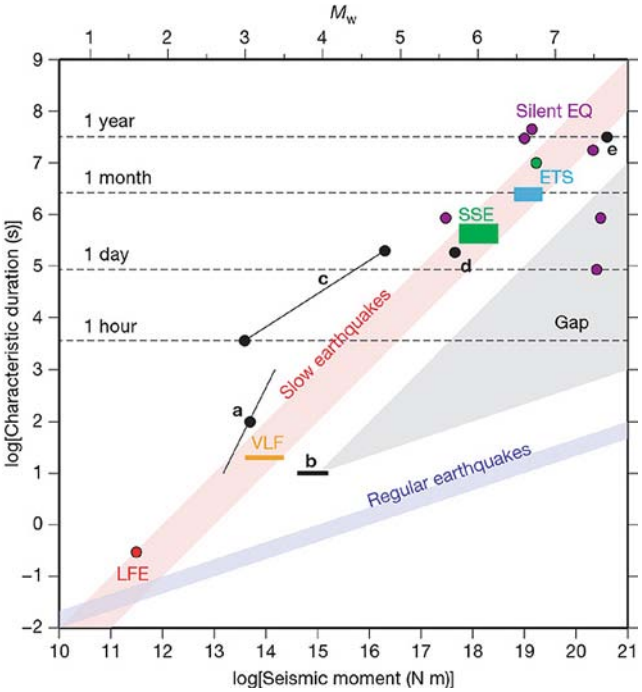
Mechanism of slow earthquake

The mechanism of fault slip has been extensively studied in the field of rock mechanics including theoretical treatments. It is inevitable to clarify the constitutive relations of fault ruptures to understand what mechanism controls the occurrence of slow earthquake (or earthquake itself). Numerous experimental and theoretical studies have been conducted for this purpose in decades. Among various proposed constitutive laws, the rate- and state-dependent friction law has been widely accepted (Dieterich, 1979; Ruina, 1983). The law is expressed by the following equations:

$$\mu\left(=\frac{\tau}{\sigma}\right)=\mu^{-}+a\ln\left(\frac{V}{V^{*}}+\theta\right) \tag{3}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}t}=-\left(\frac{V}{L}\right)\left[\theta+b\ln\left(\frac{V}{V^{*}}\right)\right] \tag{4}$$

Where, μ is friction coefficient, τ is shear stress, σ is normal stress, μ^{-} is friction coefficient at V^{*} where V^{*} is constant, L is characteristic slip distance, and θ is so-called state variable that represents the state along the fault



Slow Earthquake, Figure 7 Comparison between seismic moment and the characteristic duration of various slow earthquakes. LFE (*red*), VLF (*orange*), SSE (*green*), and ETS (*light blue*) are low-frequency earthquake, very low frequency earthquake, slow slip event, and episodic tremor and slip, respectively. See Ide et al. (2007) for more details.

surface. Coefficients a and b are constant. When the state θ does not change with time, it is called steady state and the friction coefficient in such steady state μ^{ss} is shown as

$$\mu^{ss} = \mu_0 + (a - b) \ln\left(\frac{V}{V^-}\right). \quad (5)$$

A simple spring-slider model is often used to model the fault slips (Ruina, 1983; Rice and Ruina, 1983). The system is characterized by the normal stress, shear stress, and the stiffness (or spring constant) k. Studies suggest that the type of slip, either rapid or slow slip, is controlled by several parameters. From Equation 5, it is suggested that, if $(a - b)$ is negative, the friction coefficient μ_{ss} becomes smaller when slip velocity V increases, which is called velocity weakening. In case of velocity weakening, system stiffness k controls the slip mode; if k is smaller than a critical value k_c which is given by $(b - a)\ \sigma/L$ (e.g., Dieterich, 1979), then the slip is unstable. Even in case of velocity weakening, the slip becomes slow if $(a - b)$ is close to zero. When $(a - b) > 0$, it is called velocity strengthening and the slip is decelerated.

Marone et al. (1991) used the case of velocity strengthening to interpret the post-seismic transient slow slip along the San Andreas fault and estimated the thickness of surface velocity strengthening layer to be 2–5 km. Yoshida and Kato (2003) used two tethered spring-sliders model to investigate fault interactions and examined what kind of conditions controls the occurrence of slow slip events. They showed that the repeated slow slip events may occur if the stiffness of the first block k is smaller than k_c and the second block is such that k is slightly bigger than k_c (which means that the condition is close to the boundary between stable and unstable slip). They also examined the case that an earthquake triggers the slow slip events in the adjacent region as was observed in the case of 2003 Tokachi-oki earthquake.

Further application of the rate- and state-dependent friction law to the numerical simulation of slip along the fault has been developed by a number of researchers. Tse and Rice (1986), for example, applied the relation to the two-dimensional San Andreas fault and suggested that unstable sliding may be limited to the depth shallower than 5–7 km and the steady slow slips prevails below 13–15 km of depth. Three-dimensional application was, for example, done by Rice (1993) and Stuart and Tullis (1995).

Concluding remarks

Although the slow earthquake has been studied for decades, its rapid progress was realized since the advent of the Global Positioning System observations in the middle of 1990s. Slow slip events would be one of the most important discoveries in seismology in the recent decades, thanks to GPS. In addition, the discoveries of other low-frequency events by high-sensitivity seismic arrays provided another important progress in understanding the mechanism of earthquake generation.

It may be reminded that the whole earthquake cycle was once categorized into four stages, namely, inter-seismic, pre-seismic, co-seismic, and post-seismic stages, and they had been studied rather independently. However, these stages are now recognized as cross-sectional views of single continuous series of progress of slips along the fault under the unified governing fault constitutive law.

Synthetic image of earthquake from high frequency that has been recorded by seismometers to very low frequency including transient slow events provided seismologists with opportunities of ultimate understanding of what "earthquake" is. In order, however, to realize this, a lot of more advanced researches are to be conducted and the results should be synthesized; for example, detailed deep sounding of structures of the lower crust and upper mantle where brittle and ductile ruptures are taking place should be conducted to clarify the field of earthquakes. Also, laboratory experiments for developing more conclusive governing laws of rapid and slow ruptures are indispensable. In addition, further findings of slow slip events and slow earthquakes at various fields in the world will provide deeper insights on the mechanism of earthquake generation. Such progresses might lead to a further development toward the most difficult seismological problem of earthquake prediction.

Acknowledgments

The author is indebted to Professor Naoyuki Kato and Professor Kazushige Obara of Earthquake Research Institute, the University of Tokyo, for their helpful discussions and providing information to the author. This study was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, under its Observation and Research Program for Prediction of Earthquakes and Volcanic Eruptions.

Bibliography

Amoruso, A., Crescentini, L., Morelli, A., and Scarpa, R., 2002. Slow rupture of an aseismic fault in a seismogenic region of Central Italy. *Geophys Res Lett*, **29**(24), 2219, doi:10.1029/2002GL016027.

Asano, Y., Obara, K., and Ito, Y., 2008. Spatiotemporal distribution of very-low frequency earthquakes in Tokachi-oki near the junction of the Kuril and Japan trenches revealed by using array signal processing. *Earth, Planets and Space*, **60**, 871–875.

Beroza, G. C., and Jordan, T., 1990. Searching for slow and silent earthquakes. *Journal of Geophysical Research*, **95**, 2485–2510.

Brudzinski, M., Cabral-Cano, E., Correa-Mora, F., DeMets, C., and Márquez-Azúa, B., 2007. Slow slip transients along the Oaxaca subduction segment from 1993 to 2007. *Geophysical Journal International*, **171**, 523–538.

Correa-Mora, F., DeMets, C., Cabral-Cano, E., and Márquez-Azúa, B., 2008. Interplate coupling and transient slip along the subduction interface beneath Oaxaca, Mexico. *Geophysical Journal International*, **175**, 269–290.

Crescentini, L., Amoruso, A., and Scarpa, R., 1999. Constraints on slow earthquake dynamics from a swarm in central Italy. *Science*, **286**, 2132–2134.

Davis, E. E., Becker, K., Wang, K., Obara, K., Ito, Y., and Kinoshita, M., 2006. A discrete episode of seismic and aseismic deformation of the Nankai trough subduction zone accretionary prism and incoming Philippine Sea plate. *Earth and Planetary Science Letters*, **242**, 73–84.

Dieterich, J. H., 1979. Modeling of rock friction, 1. Experimental results and constitutive equations. *Journal of Geophysical Research*, **84**, 2161–2168.

Douglas, A., Beavan, J., Wallace, L., and Townend, J., 2005. Slow slip on the northern Hikurangi subduction interface, New Zealand. *Geophysical Research Letters*, **32**, L16305, doi:10.1029/2005GL023607.

Dragart, H., Wang, K., and James, T. S., 2001. A silent slip event on the deeper Cascadia subduction interface. *Science*, **292**, 1525–1528.

Dragart, H., Wang, K., and Rogers, G., 2004. Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth Planets Space*, **56**, 1143–1150.

Fitch, T. J., and Scholz, C. H., 1971. Mechanism of underthrusting in southwest Japan: a model of convergent plate interactions. *Journal of Geophysical Research*, **76**, 7260–7292.

Gladwin, M. T., Gwyther, R. L., Hart, R. H. G., and Brechenridge, K. S., 1994. Measurements of the strain field associated with episodic creep events on the San Andreas fault at San Juan Bautista, California. *Journal of Geophysical Research*, **99**, 4559–4564.

Heki, K., Miyazaki, S., and Tsuji, H., 1997. Silent fault slip following an interpolate thrust earthquake at the Japan Trench. *Nature*, **386**, 595–598.

Hirose, H., and Obara, K., 2005. Repeating short- and long-term slow slip events with deep tremor activity around the Bungo Channel region, southwest Japan. *Earth, Planets and Space*, **57**, 961–972.

Hirose, H., and Obara, K., 2006. Short-term slow slip and correlated tremor episodes in the Tokai region, central Japan. *Geophysical Research Letters*, **33**, L17311, doi:10.1029/2006GL026579.

Hirose, H., Hirahara, K., Kimata, F., Fujii, N., and Miyazaki, S., 1999. A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo Channel, southwest Japan. *Geophysical Research Letters*, **26**(21), 3237–3240.

Ide, S., Beroza, G. C., Shelly, D. R., and Uchide, T., 2007. A scaling law for slow earthquakes. *Nature*, **447**, 76–79, doi:10.1038/nature05780.

Ishihara, Y., 2003. Major existence of very low frequency earthquakes in background seismicity along subduction zone of Southwestern Japan, *Eos Transactions of AGU*, **84** (46), Fall Meet. Suppl., Abstract S41C-0107.

Ito, Y., and Obara, K., 2006a. Dynamic deformation of the accretionary prism excites very low frequency earthquakes. *Geophysical Research Letters*, **33**, L02311, doi:10.1029/2005GL025270.

Ito, Y., and Obara, K., 2006b. Very low frequency earthquakes within accretionary prisms are very low stress-drop earthquakes. *Geophysical Research Letters*, **33**, L09302, doi:10.1029/2006GL025883.

Kanamori, H., 1972. Mechanism of tsunami earthquakes. *Physics of the Earth and Planetry Interiors*, **6**, 346–359.

Kanamori, H., and Cipar, J., 1974. Focal process of the great Chilean earthquake May 22, 1960. *Physics of the Earth and Planetary Interiors*, **9**, 127–136.

Kanamori, H., and Kikuchi, M., 1993. The 1992 Nicaragua earthquake: a slow tsunami earthquake associated with subducted sediments. *Nature*, **361**, 714–716.

Kanamori, H., and Stewart, G. S., 1979. A slow earthquake. *Physics of the Earth and Planetary Interiors*, **18**, 167–175.

Kawasaki, I., Asai, Y., Tamura, Y., Sagiya, T., Mikami, N., Okada, Y., Sakata, M., and Kasahara, M., 1995. The 1992 Sanriku-Oki, Japan, ultra-slow earthquake. *Journal of Physics of Earth*, **43**, 105–116.

Kato, A., Iidaka, T., Ikuta, R., Yoshida, Y., Katsumata, K., Iwasaki, T., Sakai, S., Thurber, C., Tsumura, N., Yamaoka, K., Watanabe, T., Kunitomo, T., Yamazaki, F., Okubo, M., Suzuki, S., and Hirata, N., 2010. Variations of fluid pressure within the subducting oceanic crust and slow earthquakes. *Geophys Res Lett*, **37**, L14310, doi:10.1029/2010GL043723.

Kawasaki, I., Asai, Y., and Tamura, Y., 2001. Space-time distribution of interplate moment release including slow earthquakes and the seismo-geodetic coupling in the Sanriku-oki region along the Japan trench. *Tectonophysics*, **330**, 267–283.

Kostoglodov, V., Singh, S. K., Santiago, J. A., Franco, S. I., Larson, K. M., Lowry, A. R., and Bilham, R., 2003. A large silent earthquake in the Guerrero seismic gap, Mexico. Geophysical Research Letters, **30**(15), 1807, doi: 10.1029/2003GL017219

Linde, A. T., and Silver, P. G., 1989. Elevation changes and the great 1960 Chilean earthquake: support for aseismic slip. *Geophysical Research Letters*, **16**, 1305–1308.

Linde, A. T., Gladwin, M. T., Johnston, M. J. S., Gwyther, R. L., and Bilham, R. G., 1996. A slow earthquake sequence on the San Andreas Fault. *Nature*, **383**, 65–68.

Linde, A. T., and Sacks, I. S., 2002. Slow earthquakes and great earthquakes along the Nankai trough. *Earth Planet Sci Lett*, **203**, 265–275.

Lowry, A. R., Larson, K. M., Kostoglodov, V., and Bilham, R., 2001. Transient fault slip in Guerrero, southern Mexico. *Geophysical Research Letters*, **28**, 3753–3756.

Marone, C. J., Scholz, C. H., and Bilham, R., 1991. On the mechanism of earthquake afterslip. *Journal of Geophysical Research*, **96**, 8441–8452.

Matsubara, M., Obara, K., and Kasahara, K., 2009. High-Vp/Vs zone accompanying non-volcanic tremors and slow-slip events beneath southwestern Japan. *Tectonophysics*, **472**, 6–17.

Melbourne, T. I., Webb, F. H., Stock, J. M., and Reigber, C., 2002. Rapid postseismic transients in subduction zones from continuous GPS. *Journal of Geophysical Research*, **107**(B10), 2241, doi:10.1029/2001JB000555.
Miller, M. M., Melbourne, T., Johnson, D. J., and Sumner, W. Q., 2002. Periodic slow earthquakes from the Cascadia subduction zone. *Science*, **295**, 2423.
Miyazaki, S., Segall, P., Fukuda, J., and Kato, T., 2004. Space time distribution of afterslip following the 2003 Tokachi-oki earthquake: implications for variations in fault zone frictional properties. *Geophysical Research Letters*, **31**, L06623, doi:10.1029/2003GL019410.
Miyazaki, S., Segall, P., McGuire, J. J., Kato, T., and Hatanaka, Y., 2006. Spatial and temporal evolution of stress and slip rate during the 2000 Tokai slow earthquake. *Journal of Geophysical Research*, **111**, B03409, doi:10.1029/2004JB003426.
Obara, K., 2002. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science*, **296**, 1679–1681.
Obara, K., and Hirose, H., 2006. Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics*, **417**, 33–51.
Obara, K., and Ito, Y., 2005. Very low frequency earthquakes excited by the 2004 off the Kii peninsula earthquakes: a dynamic deformation process in the large accretionary prism. *Earth, Planets and Space*, **57**, 321–326.
Ohta, Y., Freymueller, J. T., Hreinsdottir, Sigrun, and Suito, H., 2006. A large slow slip event and the depth of the seismogenic zone in the south central Alaska subduction zone. *Earth and Planetary Science Letters*, **247**, 108–116.
Okada, A., and Nagata, T., 1953. Land deformation of the neighborhood of Muroto point after the Nankaido great earthquake in 1946. *Bulletin of Earthquake Research Institute*, **31**, 169–177.
Ozawa, S., Murakami, M., Kaidzu, M., Tada, T., Hatanaka, Y., Yarai, H., and Nishimura, T., 2002. Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan. *Science*, **298**, 1009–1012.
Ozawa, S., Miyazaki, S., Hatanaka, Y., Imakiire, T., Kaidzu, M., and Murakami, M., 2003. Characteristic silent earthquakes in the eastern part of the Boso peninsula, Central Japan. *Geophys Res Lett*, **30**(6), 1283, doi:10.1029/2002GL016665.
Protti, M., Gonzales, V., Kato, T., Iinuma, T., Miyazaki, S., Obana, K., Kaneda, Y., LaFemina, P., Dixon, T., and Schwartz, S., 2004. A creep event on the shallow interface of the Nicoya Peninsula, Costa Ricaseismogenic zone. Eos Transactions of AGU, **85**(47), Fall Meet. Suppl., Abstract S441D-07.
Rice, J. R., 1993. Spatio-temporal complexity of slip on a fault. *Journal of Geophysical Research*, **98**, 9885–9907.
Rice, J. R., and Ruina, A. L., 1983. Stability of steady frictional slipping. *Journal of Applied Mechanics*, **50**, 343–349.
Roeloffs, E. A., 2006. Evidence for aseismic deformation rate changes prior to earthquakes. *Annual Review of Earth and Planetary Sciences*, **34**, 591–627.
Rogers, G., and Dragart, H., 2003. Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip. *Science*, **300**, 1942–1943.
Ruina, A. L., 1983. Slip instability and state variable friction laws. *Journal of Geophysical Research*, **88**, 10359–10370.
Sagiya, T., 2004. Interplate coupling in the Kanto district, Central Japan, and the Boso Peninsula silent earthquake in May 1996. *Pure and Applied Geophysics*, **161**, 2327–2342.
Schwartz, S. Y., 2009. Episodic aseismic slip at plate boundaries. In Kanamori, H. (ed.), Schubert G. (ed.-in-chief), *Earthquake Seismology*. Oxford: Elsevier, Vol. 4, pp. 445–472.
Schwartz, S. Y., and Rokosky, J. M., 2007. Slow slip events and seismic tremor at circum-Pacific subduction zones. *Review of Geophysics*, **45**, RG3004.
Segall, P., and Matthews, M., 1997. Time dependent inversion of geodetic data. *Journal of Geophysical Research*, **102**(B10), 22391–22409.
Shelly, D. R., Beroza, G. C., Ide, S., and Nakamura, S., 2006. Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature*, **442**, 188–191.
Stuart, W. D., and Tullis, T. E., 1995. Fault model for preseismic deformation at Parkfield, California. *Journal of Geophysical Research*, **100**, 24079–24099.
Tse, S. T., and Rice, J. R., 1986. Crustal earthquake instability in relation to the depth variation of frictional slip properties. *Journal of Geophysical Research*, **91**, 9452–9472.
Wesson, R. L., 1987. Modeling aftershock migration and afterslip of the San Juan Bautista, California, earthquake of October 3, 1972. *Tectonophysics*, **144**, 215–229.
Yagi, Y., and Kikuchi, M., 2003. Partitioning between co-seismic slip and asesmic slip. *Journal of Geography*, **112**(6), 828–836 (in Japanese with English abstract).
Yagi, Y., Kikuchi, M., and Sagiya, T., 2001. Co-seismic slip, post-seismic slip, and aftershocks associated with two large earthquakes in 1996 in Hyuga-nada, Japan. *Earth, Planets and Space*, **53**, 793–803.
Yagi, Y., Kikuchi, M., and Nishimura, T., 2003. Co-sesmic slip, post-seismic slip, and largest aftershock associated with the 1994 Sanriku-haruka-oki, Japan, earthquake. *Geophysical Research Letters*, **30**(22), 2177, doi:10.1029/2003GL018189.
Yoshida, S., and Kato, N., 2003. Episodic aseismic slip in a two-degree-of-freedom block-spring model. *Geophysical Research Letters*, **30**(13), 1681, doi:10.1029GL017439.
Yoshioka, S., Mikumo, T., Kostoglodov, V., Larson, K. M., Lowry, A. R., and Singh, S. K., 2004. Interplate coupling and a recent aseismic slow slip event in the Guerrero seismic gap of the Mexican subduction zone, as deduced from GPS data inversion using a Baysian information criterion. *Physics of the Earth and Planetary Interiors*, **146**, 513–530.

Cross-references

Earthquakes, Source Theory
GPS, Data Acquisition and Analysis
GPS, Tectonic Geodesy
Seismic Quiescence and Activation

SPHERICAL HARMONIC ANALYSIS APPLIED TO POTENTIAL FIELDS*

Nikolaos K. Pavlis
Geodesy and Geophysics Basic and Applied Research, National Geospatial-Intelligence Agency (NGA), Reston, VA, USA

Synonyms

Fourier expansion on the sphere; Spherical harmonic decomposition; Spherical harmonic expansion; Spherical harmonic transform

Definition

A square integrable function, defined on a surface that has a one-to-one correspondence with the unit sphere, may be

represented as a linear combination of *Surface Spherical Harmonic functions*. *Spherical Harmonic Analysis* (SHA) is the process by which the coefficients defining this linear combination are determined. These coefficients constitute the *Surface Spherical Harmonic spectrum* of the function. Functions that satisfy Laplace's partial differential equation are called *harmonic*. These can be decomposed into series of *Solid Spherical Harmonic functions*. Although SHA is *not* restricted to functions that represent potential fields, it plays a special role in the determination of such fields through the solution of *Boundary Value Problems* (BVP).

Introduction: basic formulas

A scalar function V, representing a potential field such as the Earth's gravitational or magnetic field, satisfies Laplace's partial differential equation (PDE), in the space outside of the sources generating the field. In geocentric spherical polar coordinates (r, θ, λ), where r is the radial distance, θ is the geocentric colatitude (defined as 90° minus the geocentric latitude), and λ is the longitude, Laplace's PDE takes the form (Heiskanen and Moritz, 1967, Equation 1–41):

$$\Delta V \equiv \nabla^2 V = \frac{\partial^2 V}{\partial r^2} + \frac{2}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial V}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 V}{\partial \lambda^2} = 0. \tag{1}$$

Equation 1 may be solved using the method of separation of variables. For the space outside a sphere of radius $r = a$, its solution may be written in the form (see Heiskanen and Moritz, 1967 for a complete derivation):

$$V(r,\theta,\lambda) = \sum_{n=0}^{\infty}\left(\frac{a}{r}\right)^{n+1}\sum_{m=-n}^{n} v_{nm} Y_{nm}(\theta,\lambda). \tag{2}$$

The functions $\left[(a/r)^{n+1} Y_{nm}(\theta,\lambda)\right]$ are called *Solid Spherical Harmonic functions*. The *Surface Spherical Harmonic functions* $Y_{nm}(\theta,\lambda)$ are defined as:

$$Y_{nm}(\theta,\lambda) = \overline{P}_{n|m|}(\cos\theta)\cdot\begin{Bmatrix}\cos m\lambda \\ \sin|m|\lambda\end{Bmatrix}\begin{matrix}\text{if } m \geq 0 \\ \text{if } m < 0\end{matrix}. \tag{3}$$

$\overline{P}_{n|m|}(\cos\theta)$ is the fully normalized *Associated Legendre function* of the first kind, of degree n and order $|m|$ (Heiskanen and Moritz, 1967, Sects. 1–11 and 1–14). *Fully normalized* Associated Legendre functions are commonly used in geodesy. These are related to their unnormalized counterparts $P_{n|m|}(\cos\theta)$ by:

$$\overline{P}_{n|m|}(\cos\theta) = \sqrt{(2-\delta_{|m|0})(2n+1)\frac{(n-|m|)!}{(n+|m|)!}}\cdot P_{n|m|}(\cos\theta), \tag{4}$$

where:

$$\delta_{|m|0} = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}. \tag{5}$$

For the surface spherical harmonic functions Y_{nm}, this normalization implies that:

$$\frac{1}{4\pi}\iint_{\sigma} Y_{nm}^2(\theta,\lambda)\,d\sigma = 1, \tag{6}$$

with the integration being performed over the unit sphere σ, whose area element is $d\sigma$ ($d\sigma = \sin\theta d\theta d\lambda$). Note that in geomagnetism, the *Schmidt seminormalized* Associated Legendre functions (Blakely, 1995, p. 113) $\breve{P}_{n|m|}(\cos\theta)$ are used instead. These are defined by:

$$\breve{P}_{n|m|}(\cos\theta) = \sqrt{(2-\delta_{|m|0})\frac{(n-|m|)!}{(n+|m|)!}}\cdot P_{n|m|}(\cos\theta). \tag{7}$$

Surface spherical harmonics constitute a set of *orthogonal* basis functions on the unit sphere, i.e.:

$$\iint_{\sigma} Y_{nm}(\theta,\lambda) Y_{sr}(\theta,\lambda)\,d\sigma = 0 \quad \text{if } n \neq s \text{ or } m \neq r \text{ or both.} \tag{8}$$

This property of *orthogonality* permits the determination of the spherical harmonic coefficients v_{nm}, which appear in Equation 2, as follows. Assume for the sake of this discussion that the function $V(r,\theta,\lambda)$ is observable everywhere on the surface of the sphere $r = a$. Equation 2 then takes the form:

$$V(a,\theta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} v_{nm} Y_{nm}(\theta,\lambda). \tag{9}$$

In this case, the orthogonality of surface spherical harmonics implies that:

$$v_{nm} = \frac{1}{4\pi}\iint_{\sigma} V(a,\theta,\lambda) Y_{nm}(\theta,\lambda)\,d\sigma. \tag{10}$$

The coefficients v_{nm} constitute the *surface spherical harmonic spectrum* of the function $V(r,\theta,\lambda)$, pertinent to the surface of the sphere of radius $r = a$. It should be emphasized that *spherical harmonic expansions*, as the one given in Equation 9, are *not* restricted to functions satisfying Laplace's equation (*harmonic functions*). Any square integrable function $f(\theta,\lambda)$, defined over a surface S that has a one-to-one correspondence with the unit sphere, may be expanded in surface spherical harmonics as:

$$f^S(\theta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_{nm}^S Y_{nm}(\theta,\lambda), \tag{11}$$

with the coefficients f_{nm}^S given by:

$$f_{nm}^S = \frac{1}{4\pi} \iint_\sigma f^S(\theta,\lambda) Y_{nm}(\theta,\lambda) d\sigma. \tag{12}$$

The superscript "S" emphasizes the dependence of the spectrum f_{nm}^S of the function $f(\theta,\lambda)$, on the particular surface S over which its values $f^S(\theta,\lambda)$ are given. Jekeli (1988) carefully distinguishes between the direct Legendre transform (i.e., the spectrum) of an arbitrary square integrable function $f(\theta,\lambda)$, defined on a surface S that has a one-to-one correspondence with the unit sphere, and the special case of the solution of Laplace's equation when the boundary data reside on the surface of the sphere $r = a$. The former is expressed in Equations 11 and 12 above; the latter in Equations 9 and 10. Obviously the two quantities take the same form with an appropriate choice of coordinates. Furthermore, note that while rather arbitrary functions can be expanded in surface spherical harmonics, *only* harmonic functions (i.e., functions satisfying Laplace's equation) can be expanded into solid spherical harmonics, within their region of harmonicity.

The total power of the function $f(\theta,\lambda)$, defined over the surface S, is defined to be:

$$M\left\{\left[f^S\right]^2\right\} = \frac{1}{4\pi} \iint_\sigma \left[f^S(\theta,\lambda)\right]^2 d\sigma. \tag{13}$$

A generalization of Parseval's theorem permits the computation of $M\left\{\left[f^S\right]^2\right\}$ from the spectrum f_{nm}^S, by:

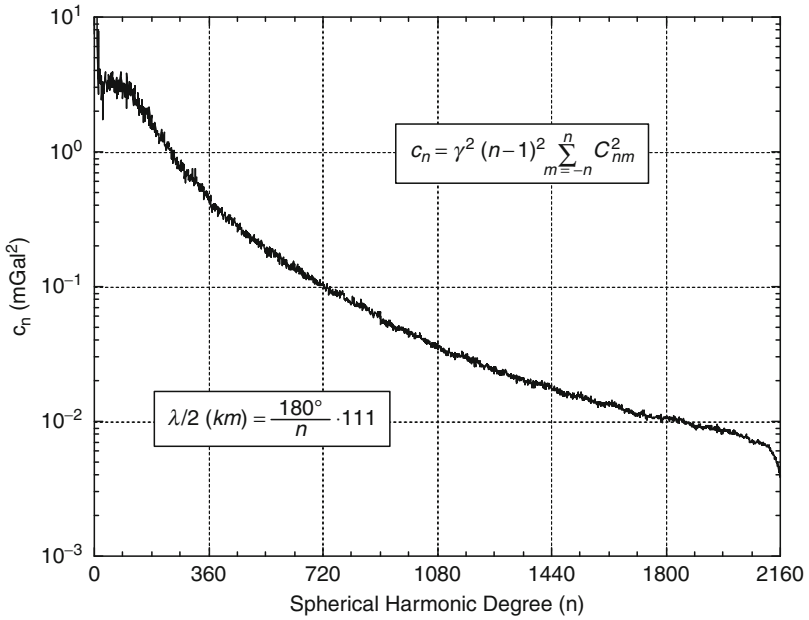
$$M\left\{\left[f^S\right]^2\right\} = \sum_{n=0}^{\infty} f_n^S = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left[f_{nm}^S\right]^2. \tag{14}$$

The quantities f_n^S, defined as:

$$f_n^S = \sum_{m=-n}^{n} \left[f_{nm}^S\right]^2, \tag{15}$$

represent the total power (or variance) per spherical harmonic degree. In geodetic literature, they are known as *degree variances*. A direct correspondence exists between spherical harmonic degree and spatial wavelength. Therefore, the degree variances enable the study of the distribution of power within the function $f^S(\theta,\lambda)$, in terms of spatial wavelength. For example, Figure 1 (Pavlis et al., 2008) shows the degree variances implied by the Earth Gravitational Model 2008 (EGM2008) for the *free-air gravity anomaly* signal.

Spherical harmonic expansions of functions, such as the Earth's topography (heights above and depths below the Mean Sea Level), surface temperature, atmospheric pressure, etc., are very useful since they permit the study of the spectral properties of these functions. Furthermore, the decomposition of these functions into their spectral components permits the efficient application of *convolutions* on the sphere (Driscoll and Healy, 1994), and of



Spherical Harmonic Analysis Applied to Potential Fields, Figure 1 Degree variances (c_n) of the free-air gravity anomaly signal (in mGal^2) implied by the EGM2008 global gravitational model, as a function of spherical harmonic degree n. The degree variances refer to the surface of a sphere of radius $a = 6,378,136.3$ m. The relationship between half-wavelength $(\lambda/2)$ resolution and spherical harmonic degree is provided in the *insert*. In SI units, $1\ \mathrm{mGal} = 10^{-5}\ \mathrm{m\,s^{-2}}$.

band-pass filtering of data, with strict control on the filter's spherical harmonic-degree bandwidth. For functions representing potential fields, in addition to these spectral analysis applications, spherical harmonic analysis offers one approach for the solution of the underlying *Boundary Value Problem* that is associated with the specific *functional* of the field that has been observed. We discuss these aspects next.

Boundary Value Problems (BVP) of potential theory

In general, a BVP consists of a differential equation that is subject to a set of boundary conditions. In potential theory, the unknown (scalar) function to be determined is the potential, V, which satisfies Laplace's PDE, in the space outside of its generating sources. The function V should vanish at infinity as the reciprocal of the distance between the point in question and the generating source element. The boundary conditions, in order to be of practical use, should represent quantities related to V (*field functionals*) that are observable over some surface S. Two general approaches exist for the solution of a BVP:

(a) Solution using spectral analysis, i.e., spherical harmonics
(b) Solution using *Greens functions*, i.e., integral formulas

Traditionally, in gravimetric geodesy, the following three BVP of potential theory have been considered:

1. First BVP of potential theory, or *Dirichlet*'s problem (Heiskanen and Moritz, 1967, Sects. 1–16): given an arbitrary function f on a surface S, determine a function V that is harmonic either inside or outside S and which assumes on S the values of the prescribed function f. If the surface S is a sphere of radius $r = a$, the solution in terms of spherical harmonics, for the region outside S, is given by Equation 2. The corresponding Greens function solution is given by *Poisson*'s integral formula (Heiskanen and Moritz, 1967, Equation 1–89).
2. Second BVP of potential theory, or *Neumann*'s problem (Heiskanen and Moritz, 1967, Sects. 1–17): here, instead of the values of the potential V itself, one is given on a surface S the values of its normal derivative $\partial V/\partial n$. The normal derivative is the derivative along the outward-directed surface normal n to S. If the surface S is a sphere of radius $r = a$, the solution in terms of spherical harmonics, for the region outside S, is given by:

$$V(r,\theta,\lambda) = -a\sum_{n=0}^{\infty}\frac{1}{n+1}\left(\frac{a}{r}\right)^{n+1}\sum_{m=-n}^{n}u_{nm}Y_{nm}(\theta,\lambda), \tag{16}$$

where u_{nm} is the spectrum of the $\partial V/\partial n$ boundary values, given on the sphere $r = a$, i.e.:

$$\left(\frac{\partial V}{\partial n}\right)_{r=a} = \left(\frac{\partial V}{\partial r}\right)_{r=a} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}u_{nm}Y_{nm}(\theta,\lambda). \tag{17}$$

The values of u_{nm} can be obtained from:

$$u_{nm} = \frac{1}{4\pi}\iint_{\sigma}\left(\frac{\partial V}{\partial r}\right)_{r=a}Y_{nm}(\theta,\lambda)d\sigma. \tag{18}$$

The corresponding Greens function solution to the second BVP is given by *Hotine*'s integral formula (Hotine, 1969).

3. Third BVP of potential theory (Heiskanen and Moritz, 1967, Sects. 1–17): here, a linear combination of V and of its normal derivative $\partial V/\partial n$ is given on the surface S, i.e., $hV + k(\partial V/\partial n)$ is given on S. Again, if the surface S is a sphere of radius $r = a$, the solution in terms of spherical harmonics, for the region outside S, is given by:

$$V(r,\theta,\lambda) = \sum_{n=0}^{\infty}\frac{1}{h-(k/a)(n+1)}\left(\frac{a}{r}\right)^{n+1}\sum_{m=-n}^{n}w_{nm}Y_{nm}(\theta,\lambda), \tag{19}$$

where w_{nm} is the spectrum of the $hV + k(\partial V/\partial n)$ boundary values, given on the sphere $r = a$, i.e.:

$$\left(hV + k\frac{\partial V}{\partial n}\right)_{r=a} = \left(hV + k\frac{\partial V}{\partial r}\right)_{r=a} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}w_{nm}Y_{nm}(\theta,\lambda). \tag{20}$$

The values of w_{nm} can be obtained from:

$$w_{nm} = \frac{1}{4\pi}\iint_{\sigma}\left(hV + k\frac{\partial V}{\partial r}\right)_{r=a}Y_{nm}(\theta,\lambda)d\sigma. \tag{21}$$

This particular BVP is of great importance to physical geodesy. The determination of the *geoid* given *gravity anomalies* as boundary data – called the BVP of physical geodesy – represents a specific case of this third BVP, where $h = -2/a$ and $k = -1$. With these values for h and k, if the surface S is a sphere of radius $r = a$, the solution in terms of spherical harmonics, for the region outside S, is given by:

$$V(r,\theta,\lambda) = a\sum_{n=0}^{\infty}\frac{1}{n-1}\left(\frac{a}{r}\right)^{n+1}\sum_{m=-n}^{n}g_{nm}Y_{nm}(\theta,\lambda), \tag{22}$$

where g_{nm} is the spectrum generated from the SHA of the gravity anomaly boundary values that are given on the sphere $r = a$. The corresponding Greens

function solution to the BVP of physical geodesy is given by *Stokes'* integral formula (Heiskanen and Moritz, 1967, Sects. 2–16).

Other BVP may also be formulated and solved, which correspond to other observable boundary data. For example, one may consider the *tensor* of second-order gradients of the gravitational potential (Moritz, 1980, p. 4) (or some subset of the elements of this tensor) as boundary data, which give rise to BVP corresponding to *gravity gradiometry*.

Equations 9, 17, and 20 are all of the same form as Equation 11. Therefore, provided that the boundary data are given on, or can be reduced to, the surface of a sphere of radius $r = a$, surface SHA may be used to determine the surface spherical harmonic spectrum of the field. This surface spectrum, multiplied by the appropriate radial terms that appear in Equations 2, 16, and 19, provides then a solution in terms of solid spherical harmonics, to the corresponding BVP. The convergence of the series 2, 16, and 19 is guaranteed in the space outside of the sphere $r = a$.

Laplace's PDE can also be expressed in terms of the *ellipsoidal coordinates* (u, δ, λ), where u is the semi-minor axis of the confocal ellipsoid, δ is the *reduced* colatitude (defined as 90° minus the reduced latitude), and λ is the longitude (Heiskanen and Moritz, 1967, Sects. 1–19). The Cartesian coordinates (x, y, z) of a point are related to its spherical (r, θ, λ) and ellipsoidal (u, δ, λ) coordinates as follows:

$$\begin{aligned} x &= r\sin\theta\cos\lambda = \left(u^2 + E^2\right)^{1/2}\sin\delta\cos\lambda \\ y &= r\sin\theta\sin\lambda = \left(u^2 + E^2\right)^{1/2}\sin\delta\sin\lambda, \\ z &= r\cos\theta = u\cos\delta \end{aligned} \tag{23}$$

where E is the constant linear eccentricity of the family of confocal ellipsoids defining the coordinate system. Expressed in the (u, δ, λ) coordinate system, Laplace's PDE can also be solved using the method of separation of variables. The solution, for the space exterior to a reference ellipsoid with semi-minor axis $u = b$, is given by (Heiskanen and Moritz, 1967, Sects. 1–20):

$$V(u, \delta, \lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{Q_{n|m|}(i(u/E))}{Q_{n|m|}(i(b/E))}z_{nm}Y_{nm}(\delta, \lambda). \tag{24}$$

$Q_{n|m|}(i(u/E))$ is the *Associated Legendre function* of the second kind, of degree n and order $|m|$ (Heiskanen and Moritz, 1967, Sects. 1–12). i is the imaginary unit $(i = \sqrt{-1})$. On the surface of the reference ellipsoid $u = b$, Equation 24 becomes:

$$V(b, \delta, \lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}z_{nm}Y_{nm}(\delta, \lambda). \tag{25}$$

Equation 25 is of the same general form as Equations 9 and 11, with the important distinction that δ now represents the *reduced* colatitude. The surface *ellipsoidal* harmonic spectrum z_{nm} can be determined from:

$$z_{nm} = \frac{1}{4\pi}\iint_{\sigma} V(b, \delta, \lambda)Y_{nm}(\delta, \lambda)d\sigma, \tag{26}$$

in exactly the same fashion as in the case of the surface *spherical* harmonic spectrum in 10 and 12. Note that the integration is again over the unit sphere, whose area element is now defined as $d\sigma = \sin\delta d\delta d\lambda$.

Spherical harmonics are used extensively in geodesy because of their simplicity, and because the Earth, to a first-order approximation, is a sphere. Since, to a second-order approximation, the Earth resembles more an ellipsoid of revolution, ellipsoidal harmonics are more suitable for the solution of BVP, when the boundary data reside on or near the surface of the Earth. The transformation between ellipsoidal and spherical harmonics can be performed using the exact analytical relationships formulated by Jekeli (1988) and implemented by Gleason (1988). In this fashion, the use of the rather complicated solid ellipsoidal harmonics of Equation 24 may be avoided altogether.

Spherical Harmonic Analysis (SHA): numerical techniques

Let us consider Equations 11 and 12. Equation 11 is representative of Equations 9, 17, 20, and 25, while Equation 12 is representative of Equations 10, 18, 21, and 26. Each one of Equations 11 and 12 suggests an alternative approach for the estimation of the coefficients f^S_{nm}, as we discuss next.

Least Squares Adjustment. The first approach uses the linear mathematical model of Equation 11 to set up a system of *observation equations*. With f^S_{obs} representing the observable quantity and $\hat{f}^S_{nm}$ the estimates of the unknown parameters, the observation equations can be written as:

$$\upsilon(\theta, \lambda) = \sum_{n=0}^{N}\sum_{m=-n}^{n}\hat{f}^S_{nm}Y_{nm}(\theta, \lambda) - f^S_{obs}(\theta, \lambda), \tag{27}$$

where $\upsilon(\theta, \lambda)$ is the residual associated with the observation $f^S_{obs}(\theta, \lambda)$. Notice that in Equation 27 the summation over spherical harmonic degree was truncated to a finite maximum degree N. In vector-matrix notation, Equation 27 takes the form:

$$\mathbf{v} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{l}. \tag{28}$$

$\mathbf{v}$ is the vector representing the residuals $\upsilon(\theta, \lambda)$, $\mathbf{A}$ is the *design matrix* containing the terms that multiply the unknown parameters $\hat{f}^S_{nm}$, $\hat{\mathbf{x}}$ is the vector representing the unknown parameters $\hat{f}^S_{nm}$, and $\mathbf{l}$ is the vector containing the observations $f^S_{obs}(\theta, \lambda)$. In Equation 28, only the matrix

A and the vector **l** are known quantities. The linear system (28) can be solved using the method of *Least Squares Adjustment*. If $\mathbf{\Sigma}$ denotes the variance–covariance matrix of the errors associated with the observations $f^S_{obs}(\theta,\lambda)$, then, minimization of the quadratic form:

$$\phi = \mathbf{v^T \Sigma^{-1} v}, \tag{29}$$

subject to the condition (28), leads to the *normal equations*:

$$\mathbf{N\hat{x}} = \mathbf{U}, \text{where} \quad \mathbf{N} = \mathbf{A^T \Sigma^{-1} A} \quad \text{and} \quad \mathbf{U} = \mathbf{A^T \Sigma^{-1} l}. \tag{30}$$

A scaled version of the inverse of the variance–covariance matrix is the weight matrix **P**. $\mathbf{P} = \sigma_0^2 \mathbf{\Sigma}^{-1}$ and σ_0^2 is called the a priori variance of unit weight. σ_0^2 represents the error variance of an observation whose weight is equal to 1. The solution of the system of normal equations is:

$$\hat{\mathbf{x}} = \mathbf{N}^{-1}\mathbf{U}. \tag{31}$$

The variance-covariance matrix of the errors associated with the estimated values $\hat{\mathbf{x}}$ is given by:

$$\mathbf{\Sigma}_{\hat{\mathbf{x}}} = \mathbf{N}^{-1} = \left(\mathbf{A^T \Sigma^{-1} A}\right)^{-1}. \tag{32}$$

Carl Friedrich Gauss (1777–1855), who is credited with the development of the Least Squares Adjustment technique, used this approach in the analysis of geomagnetic data. Herein, we abbreviate the Least Squares Adjustment technique by LS.

Although in the previous discussion we presented the application of the LS method to a mathematical model that describes a *surface* spherical harmonic expansion, the LS method could be applied equally well to mathematical models describing *solid* spherical harmonic expansions, like those of Equations 2, 16, and 19. This means that one may use the LS method to estimate the spectra of potential fields, even when the boundary data are arbitrarily located in the three-dimensional space. In contrast, the *Numerical Quadrature* technique that we discuss next, when used for the determination of potential fields, requires that the boundary data are located over surfaces that are coordinate surfaces in the coordinate system in which Laplace's equation is formulated and solved ($r = const.$ for spherical harmonics, $u = const.$ for ellipsoidal harmonics). In theory, these coordinate surfaces must also encompass all the sources generating the field.

Numerical Quadrature. In the above discussion of the LS technique, the distribution, both in terms of location and in terms of number, of the measured values $f^S_{obs}(\theta,\lambda)$ remained unspecified. Consider now an equiangular ($\Delta\theta = \Delta\lambda$) partition of the unit sphere, along meridians ($\lambda = const.$) and parallels ($\theta = const.$). Such a partition creates a grid of $L \times 2L$ cells on the unit sphere, where $L = \pi/\Delta\theta$. Let us assume that the function $f(\theta,\lambda)$ was sampled in such a way that, for each one of the $L \times 2L$ cells, a value $f^S_{obs}(\theta,\lambda)$ exists that corresponds, for example, to the center of that cell. Such a sampling suggests a possible *discretization* of the surface integral of Equation 12, leading to the following *Numerical Quadrature* (NQ) formula (Colombo, 1981, Equation 1.5):

$$\tilde{f}^S_{nm} = \frac{1}{4\pi} \sum_{i=0}^{L-1} \sum_{j=0}^{2L-1} f^S_{obs}(\theta_i,\lambda_j) Y_{nm}(\theta_i,\lambda_j) \Delta\sigma_i, \tag{33}$$

where the area element $\Delta\sigma_i$ of the cells residing on the ith "row" is given by:

$$\Delta\sigma_i = \Delta\lambda \int_{\theta_i}^{\theta_{i+1}} \sin\theta d\theta = \Delta\lambda \cdot (\cos\theta_i - \cos\theta_{i+1}). \tag{34}$$

Equation 33 is a simple NQ formula, applicable to data $f^S_{obs}(\theta,\lambda)$ that represent point values.

In general, the estimate of the spectrum obtained either by using LS or by using NQ, will be different from the "true" spectrum f_{nm} of the function $f(\theta,\lambda)$. Furthermore, even when the two techniques are applied to the same data $f^S_{obs}(\theta,\lambda)$, their estimates will be, in general, different, i.e.:

$$\hat{f}^S_{nm} \neq \tilde{f}^S_{nm}. \tag{35}$$

It is therefore appropriate to consider the specific errors that affect $\hat{f}^S_{nm}$ and $\tilde{f}^S_{nm}$, compare the LS and NQ techniques, and determine if it is possible to design and formulate any other SHA technique(s) that would satisfy some desirable optimality criterion. This study was essentially carried out by Colombo (1981), in his treatise of the subject, entitled *Numerical Methods for Harmonic Analysis on the Sphere*. We summarize in the following Colombo's developments and provide some examples of practical application of his formulations. Before this, however, we need to define and discuss briefly the error sources that affect any estimate of the spectrum of a function that is obtained on the basis of observed data, regardless of the technique that is used to obtain this spectrum.

- *Error Sources.* Two types of errors affect any estimate of the spectrum that is obtained from observations of a function: (a) Sampling error, and, (b) Propagated noise. The sampling error arises due to the fact that in reality, observations can only be made at a finite number of discrete points, while Equations 11 and 12 require the function $f(\theta,\lambda)$ to be known at *every* point on the surface S. Propagated noise, on the other hand, arises due to the fact that observations can never be error-free. Therefore, errors in the observations (both random and systematic) will generally propagate into the spectrum that is estimated from them.

Optimal Estimation. The particular way that sampling errors and propagated noise affect the estimated spectrum depends on the particular technique used to estimate the

spectrum. Colombo (1981) designed and formulated a linear estimator of the spectrum, which is optimal in the sense that it minimizes the sum of the squares of the sampling error and the propagated noise. His *Optimal Estimation* (OE) technique is a form of Numerical Quadrature, where the *optimal* quadrature weights are defined using the formalism of *Least Squares Collocation* (LSC) (Moritz, 1980). By exploiting the geometry of regular grids (i.e., grids where at least the longitude increment is constant) of data located over surfaces that have rotational symmetry (such as a sphere or a rotational ellipsoid), and by imposing certain conditions on the *covariance functions* of the signal and the data noise, Colombo (1981) demonstrated that the variance–covariance matrix of such data sets consists of *Toeplitz-circulant* blocks. These matrix blocks can be formed and inverted very efficiently, even when the size of the data set is very large. Colombo's OE technique was used in the gravitational field expansion developed by Hajela (1984), and in the OSU86C/D global gravitational models developed by Rapp and Cruz (1986a). Additional discussion about the estimation of gravitational spectra using LSC can be found in Tscherning (2001).

Colombo (1981, p. 76) also investigated the use of a semiempirical set of numerical quadrature weights, which depend only on the spherical harmonic degree, and can be evaluated very easily as a function of Pellinen's smoothing factors (Colombo, 1981, p. 85). Although suboptimal, these numerical quadrature weights proved to be quite adequate and were used in several gravitational modeling studies. The OSU86E/F (Rapp and Cruz, 1986b) and the OSU89A/B (Rapp and Pavlis, 1990) global gravitational models were developed using these semiempirical numerical quadrature weights.

Practical aspects

Several factors complicate the application of SHA techniques when analyzing real data. These are related on the one hand to the inherent properties of the alternative SHA techniques, and on the other to the properties of the data to be analyzed. The former have to do with the computational capabilities and requirements (i.e., the flexibility and efficiency) of the different SHA techniques. The latter include issues related to the distribution and type (point or area-mean values) of the available data, the geometry of the grid to which the original data are reduced, the spectral properties of the function $f(\theta, \lambda)$ that is being approximated, and the spectral properties of the errors that may affect the observed values $f^S_{obs}(\theta, \lambda)$. We discuss briefly these factors in the following paragraphs.

- *Computational Issues.* An inspection of Equations such as (11) indicates that a spherical harmonic expansion complete to degree and order N contains $(N+1)^2$ harmonic coefficients f^S_{nm} in total. Therefore, the normal matrix $\mathbf{N}$ of Equation 30 has dimensions $(N+1)^2 \times (N+1)^2$. The structure and characteristics of the matrix $\mathbf{N}$ depends on the geometry of the available data distribution and on the characteristics of the variance–covariance matrix $\mathbf{\Sigma}$ of the data errors. If the data to be analyzed are arbitrarily located in latitude and longitude, as is the case with most raw geophysical data collected in the field (e.g., gravimetric or magnetometric data), then, apart from symmetry, the matrix $\mathbf{N}$ would be fully occupied. The same is true (in general) if the error variance–covariance matrix $\mathbf{\Sigma}$ is arbitrary. The size of the normal matrix $\mathbf{N}$ is already formidable for expansions complete to degree and order 200 or so. Normal matrices for expansions to degree and order 2160, as the one corresponding to the EGM2008 model (Pavlis et al., 2008), are well beyond current computational capabilities. Nevertheless, provided that the size of the normal matrix can be handled computationally, LS adjustment and LSC are the *only* techniques that can handle arbitrarily located data, with arbitrary error variance–covariance matrices. NQ and OE require the data to be input in the form of a grid, and impose restrictions on the stochastic properties of the noise, represented within the matrix $\mathbf{\Sigma}$. In the following, we assume that the raw data have been preprocessed in some fashion, so that the input data to the various estimators are available in gridded form.

As we mentioned previously, unlike estimators based on orthogonality relations, LS permits also the use of *solid* spherical harmonics in the formation of observation equations. Assume for the sake of this example, that a grid of potential values V is available over the physical surface of the Earth (the topography). One may then form observation equations on the basis of the mathematical model given in Equation Equation 2. In such a case, even if the V values are given over a regular latitude-longitude grid, these values would still be located arbitrarily in the radial direction. This would then require the formation and inversion of a fully occupied normal matrix, as it is described by Pavlis (1988), who performed such estimations to degree 36, 50, and 70, using equiangular grids of free-air gravity anomalies defined on the surface of the Earth's topography.

- *Data Organization, Types, and Distribution.* In most geodetic and geophysical SHA applications, the data are organized in geographic grids. Equiangular grids are most often used, although grids with cells of equal area are also possible, and have been used in some cases. There is therefore a need to transition from the arbitrarily located point measurements, to "artificial" measurements referring to the nodes of some specified grid. In geodetic practice, this transition is accomplished by predicting what the "artificial" grid measurements would have been, given the available arbitrarily scattered data. The prediction is done using the LSC formalism, which requires some knowledge of the signal and noise covariance functions of the data (Moritz, 1980). One usually applies LSC to predict area-mean values for the grid cells, rather than point values. SHA can be applied equally well using either point or

area-mean gridded data. In the case of area-mean values, the mathematical models have to be modified appropriately, so that they accurately represent the area-mean character of the data. This is done by integrating the surface spherical harmonic functions over the specific latitudinal and longitudinal limits of each grid cell. From Equation 3 one has:

$$\int_{\theta_i}^{\theta_{i+1}} \int_{\lambda_j}^{\lambda_{j+1}} Y_{nm}(\theta,\lambda)\sin\theta d\theta d\lambda = \int_{\theta_i}^{\theta_{i+1}} \overline{P}_{n|m|}(\cos\theta)\sin\theta d\theta \int_{\lambda_j}^{\lambda_{j+1}} \begin{Bmatrix} \cos m\lambda \\ \sin|m|\lambda \end{Bmatrix} d\lambda \quad \begin{matrix} \text{if } m \geq 0 \\ \text{if } m < 0 \end{matrix}. \tag{36}$$

Efficient recurrence relationships for evaluating the integrals of the Associated Legendre functions have been derived by Paul (1978). The numerical treatment of the evaluation of Associated Legendre functions of very high degree (e.g., $n > 1,400$ or so) requires special care since some of the values of these functions become exceedingly small, which could cause numerical problems.

Even after the gridding of geophysical data collected on or near the Earth's surface, the radial location of these data, corresponding to the Earth's topography, remains arbitrary. Solutions of BVP that are based on orthogonality relations, formulated in either spherical or ellipsoidal harmonics, require these data to be located over a sphere or an ellipsoid, respectively. This artificial (approximately radial) "movement" of the data can be accomplished using *analytical continuation*. Wang (1987, 1988) discusses some techniques that can be applied to analytically continue gravity anomaly data.

In many geophysical problems where SHA techniques are applied, the available data do not suffice to cover completely the Earth. Data gaps can produce undesirable *leakage* effects in the spectral estimates, which by definition are of global character, and require global coverage. These undesirable effects are different depending on the particular technique used to estimate the spectrum. In LS, the spectral estimator "interprets" a data gap as an undefined data value with infinite error, i.e., zero weight (see Pavlis, 1988 for details). In NQ techniques, the estimator "interprets" a data gap as a data value that is equal to zero. In either case, the estimated spectrum would produce undesirable results when evaluated over the region of the data gap. It is therefore preferable to "fill in" data gaps with some reasonable values of the quantity under consideration. These "fill-in" values may be obtained on the basis of other data and/or models. For example, in gravity modeling, topographic elevations and models of the *isostatic compensation* may be used to "fill in" areas void of actual gravity measurements (cf. Pavlis and Rapp, 1990).

- *LS versus NQ techniques*. It is of interest to compare the LS and the NQ techniques from an analytical as well as a numerical perspective. Such studies have been reported by Rapp (1969, 1986), Colombo (1981), and Pavlis (1988). Sneeuw (1994) discussed the two techniques from a historical perspective. A brief review of the conclusions reached by these investigations follows. Detailed derivations and analyses supporting these conclusions can be found in the cited references. NQ is used here to identify the simple numerical quadratures formula with the semiempirical set of suboptimal quadrature weights proposed by Colombo (1981, p. 76).
  1. NQ determines each harmonic coefficient independently of all others. In contrast, LS estimates a correlated set of coefficients; thus, solutions to different maximum degrees will yield different values for the common harmonic coefficients. This will occur even if the input data are uncorrelated, and arises from the loss of orthogonality between the discrete (point or integrated) samples of the Associated Legendre functions (Pavlis, 1988, Sect. 4.2.1). Unlike the Legendre functions, discrete (point, as well as integrated) samples of $\cos m\lambda$ and $\sin m\lambda$ preserve their orthogonality in the interval $[0, 2\pi)$, as long as the sampling interval $\Delta\lambda$ is constant. It is this orthogonality property of sines and cosines, along with the equatorial symmetry and the parity properties of Legendre functions, which produce the sparsity patterns of the *Block-Diagonal* (*BD*) *Least Squares* technique that we discuss next.
  2. NQ cannot account for varying accuracies of the input data, while the LS estimator is capable of accounting for any (positive-definite) error variance–covariance matrix $\boldsymbol{\Sigma}$ associated with the input data.
  3. If $L(=\pi/\Delta\theta)$ denotes the *Nyquist degree* implied by the data sampling interval, then the normal equations formed based on observation equations like (27) become singular if $N \geq L$ (Colombo, 1981), where N is the maximum solved-for degree.
  4. LS estimation (and LSC) can recover *exactly* a set of coefficients from "synthetic" noiseless data, provided that the data are band-limited and their frequency content does not exceed the Nyquist degree. This property offers also one verification test that any software developed to perform SHA using the LS technique should pass (Pavlis, 1988, Sect. 4.2.2). The simple NQ technique is incapable of recovering the original "true" coefficients from which the "synthetic" data were computed, as Rapp's (1986) numerical experiments have also demonstrated.
  5. Unlike the LS technique, the simple NQ does not involve any matrix formation or inversion. In this

regard, the simple NQ is considerably less demanding computationally than LS. In addition, if the sampling interval $\Delta\lambda$ in the longitude direction is constant, then the NQ SHA algorithm can benefit enormously from the application of *Fast Fourier Transform* (FFT) techniques, as it was pioneered by Colombo (1981).

6. In contrast to the simple NQ, the LS technique produces not only the estimates of the coefficients, but also an estimate of their error variance–covariance matrix (see Equation 32). This matrix can then be used to propagate the errors of the estimated coefficients onto other quantities that may be computed from them.

Block-Diagonal (BD) Least Squares. From the previous discussion, it becomes clear that a SHA technique that would combine the computational efficiency of the simple NQ, with the rigor and flexibility of LS, is highly desirable. Such a technique would attempt to retain the advantages of both techniques, while avoiding their respective limitations. The Block-Diagonal (BD) Least Squares may be viewed as such a "best of both worlds" type of technique. Colombo (1981) has shown that if:

(a) The data reside on a surface of revolution (e.g., a rotational ellipsoid or a sphere)
(b) The grid is complete and the longitude increment is constant
(c) The data weights are longitude-independent
(d) The data weights are symmetric with respect to the equator

then, the elements of the normal matrix **N**, corresponding to the coefficients f^S_{nm} and f^S_{rq}, would be zero as prescribed by (see also Pavlis, 1988 for details):

$$[\mathbf{N}]_{f^S_{nm} f^S_{rq}} = 0 \quad \text{if } m \neq q, \quad \text{or if } m = q \text{ and } n - r = 2k + 1. \tag{37}$$

Note that in this notation the order subscript is a signed integer, whose sign identifies the type of coefficient (positive: cosine, negative: sine). If condition (d) does not hold true, then:

$$[\mathbf{N}]_{f^S_{nm} f^S_{rq}} = 0 \quad \text{if } m \neq q. \tag{38}$$

The sparsity patterns implied by Equations 37 and 38 will be referred to as BD1 and BD2 respectively. In addition, a BD3 pattern may be considered, defined by:

$$[\mathbf{N}]_{f^S_{nm} f^S_{rq}} = 0 \quad \text{if } |m| \neq |q|, \tag{39}$$

which admits nonzero off-diagonal elements across coefficients of different type within a given order. It is instructive to consider the computational efficiency implied by these patterns. Table 1, which is taken from Pavlis, pp. 8–5, in Lemoine et al. (1998), provides relevant statistics for an expansion complete from degree and order 0 to degree and order 360, excluding degree $n = 1$ terms. Such an expansion involves 130,318 unknown coefficients and the upper (or lower) triangular part of the symmetric but fully occupied normal matrix **N** has 8,491,455,721 elements in total.

The enormous computational savings that can be inferred from Table 1 make the BD approximations very attractive estimation techniques. These savings, however, come at the expense of the rigor exercised in the implementation of the SHA. Any approximation of a normal matrix by a BD structure should be simple enough to permit efficient numerical implementation, and, on the same time, rigorous enough to maintain the most important characteristics of the full matrix. Therefore, one has to consider carefully whether the approximations leading to the various BD sparsity patterns can be tolerated, given the characteristics of the particular data set that is analyzed. For example, in global gravitational modeling applications, the real-world gravity anomaly data to be analyzed comply *only* with the conditions (a) and (b) above. In fact, to comply even with the (a) and (b) conditions, "filling-in" techniques and analytical continuation have to be employed, since the original data set is neither complete, nor residing on any surface of revolution, since it refers to the Earth's topography. These aspects were carefully considered in the development of the OSU91A (Rapp et al., 1991) and the EGM96 (Lemoine et al., 1998) gravitational models. In both these models, different SHA techniques were used toward the estimation of different spectral portions of the models. This was due to the fact that different spectral portions of these models were determined on the basis of different gravimetric data sets.

In the case of *band-limited* data, one can show that the LS adjustment approach using a priori information is

Spherical Harmonic Analysis Applied to Potential Fields, Table 1 Statistics of normal matrices related to an expansion complete to degree $N = 360$ (excluding degree $n = 1$ coefficients) using different sparsity patterns

	Sparsity pattern		
Statistic	BD1	BD2	BD3
Total number of nonzero elements	7,905,721	15,746,100	31,362,241
Percentage of full matrix elements	0.09	0.19	0.37
Number of blocks	1,440	721	361
Number of unknowns in largest block	181	360	718
Number of elements in largest block	16,471	64,980	258,121

formally equivalent to LSC (Moritz, 1980, p. 166; Colombo, 1981, Sect. 2.13). In such a case, the BD approaches discussed previously are the Least Squares counterpart of Colombo's (1981) Optimal Estimation technique.

BD techniques of varying sophistication have been used to develop GPM2 (Wenzel, 1985), DGFI92A (Gruber and Bosch, 1992), GFZ95A (Gruber et al., 1996), the GPM98A, B, and C models (Wenzel, 1998, 1999), and the EGM2008 model (Pavlis et al., 2008).

Some of the BD structures of the normal matrix that were discussed above arise also in the analysis of data from certain configurations of *Satellite-to-Satellite Tracking* (SST) and of *Satellite Gravity Gradiometry* (SGG). This was also recognized and studied by Colombo (1984, 1989). Schuh (1996) provides a detailed study of the numerical solution strategies applicable to the analysis of such observables. SST data (in the "low-low" configuration) are currently available from the GRACE mission (GRACE, 1998), and SGG data are currently being collected from the GOCE mission (ESA, 1999). The availability of BD normal matrices resulting from the analysis of SST and SGG data is very important when considering the combination of these matrices with corresponding matrices derived from the analysis of global grids of gravimetric data. Such a combination of BD normal matrices permits the efficient development of global gravitational models to a very high degree (2160), as it was done in the case of EGM2008 (Pavlis et al., 2008).

Summary

Spherical Harmonic Analysis (SHA) is a critical element of the solution of Boundary Value Problems associated with potential fields. SHA is applicable not only to potential fields but also to a rather wide class of functions that are sampled over surfaces which have a one-to-one correspondence with the unit sphere. Least Squares (using either fully occupied or block-diagonal normal matrices), Numerical Quadrature, and Optimal Estimation techniques have been used to perform SHA within gravitational and magnetic field modeling efforts. The organization of the data to be analyzed in the form of geographic grids defined by meridians and parallels enables the use of highly efficient numerical algorithms for SHA. In particular, grids with constant longitudinal spacing of data permit the application of Fast Fourier Transform techniques, thereby increasing tremendously the efficiency of the SHA algorithms.

Bibliography

Blakely, R. J., 1995. *Potential Theory in Gravity and Magnetic Applications*. Cambridge: Cambridge University Press.

Colombo, O. L., 1981. *Numerical methods for harmonic analysis on the sphere*. Report 310, Department of Geodetic Science and Surveying, Ohio State University, Columbus.

Colombo, O. L., 1984. *The Global Mapping of Gravity with Two Satellites*. Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 7, No. 3. Delft.

Colombo, O. L., 1989. High resolution analysis of satellite gradiometry. In Rapp, R. H. (ed.), *Progress in the Determination of the Earth's Gravity Field*, Extended Abstracts from the meeting held in Ft. Lauderdale, Florida, September 13–16, 1988. Report 397, Department of Geodetic Science and Surveying, Ohio State University, Columbus.

Driscoll, J. R., and Healy, D. M., Jr., 1994. Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, **15**, 202–250.

ESA SP-1233 (1), 1999. *The Four Candidate Earth Explorer Core Missions – Gravity Field and Steady-State Ocean Circulation Mission*. Noordwijk: ESA Publications Division, ESTEC.

Gleason, D. M., 1988. Comparing ellipsoidal corrections to the transformation between the geopotential's spherical and ellipsoidal spectrums. *Manuscripta Geodaetica*, **13**, 114–129.

GRACE, 1998. – Gravity Recovery and Climate Experiment: Science and Mission Requirements Document, revision A, *JPLD-15928*, NASA's Earth System Science Pathfinder Program.

Gruber, T., and Bosch, W., 1992. *A new 360 gravity field model, presented at the XVII General*. Assembly of the European Geophysical Society, Edinburgh, Scotland.

Gruber, T., Anzenhofer, M., and Rentsch, M., 1996. The 1995 GFZ high resolution gravity model. In Rapp, R. H., Cazenave, A. A., and Nerem, R. S. (eds.), *Global Gravity Field and Its Temporal Variations*. Berlin/Heidelberg: Springer. IAG Symposia, Vol. 116.

Hajela, D. P., 1984. *Optimal estimation of high degree gravity field from a global set of 1° × 1° anomalies to degree and order 250*. Report 358, Department of Geodetic Science and Surveying, Ohio State University, Columbus.

Heiskanen, W. A., and Moritz, H., 1967. *Physical Geodesy*. San Francisco: W.H. Freeman.

Hotine, M., 1969. *Mathematical Geodesy*. Washington, DC: U.S. Department of Commerce. ESSA Monograph, Vol. 2.

Jekeli, C., 1988. The exact transformation between ellipsoidal and spherical harmonic expansions. *Manuscripta Geodaetica*, **13**, 106–113.

Lemoine, F. G., Kenyon, S. C., Factor, J. K., Trimmer, R. G., Pavlis, N. K., Chinn, D. S., Cox, C. M., Klosko, S. M., Luthcke, S. B., Torrence, M. H., Wang, Y. M., Williamson, R. G., Pavlis, E. C., Rapp, R. H., and Olson, T. R., 1998. The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96. NASA Tech. Publ. TP-1998-206861. 575 pp.

Moritz, H., 1980. *Advanced Physical Geodesy*. Karlsruhe: Herbert Wichmann.

Paul, M. K., 1978. Recurrence relations for integrals of Associated Legendre functions. *Bulletin Géodésique*, **52**, 177–190.

Pavlis, N. K., 1988. *Modeling and estimation of a low degree geopotential model from terrestrial gravity data*. Report 386, Department of Geodetic Science and Surveying, Ohio State University, Columbus.

Pavlis, N. K., and Rapp, R. H., 1990. The development of an isostatic gravitational model to degree 360 and its use in global gravity modelling. *Geophysical Journal International*, **100**, 369–378.

Pavlis, N. K., Holmes, S. A., Kenyon, S. C., and Factor, J. K., 2008. An earth gravitational model to degree 2160: EGM2008. In *Presented at the 2008 General Assembly of the European Geosciences Union*, April 13–18, Vienna, Austria.

Rapp, R. H., 1969. Analytical and numerical differences between two methods for the combination of gravimetric and satellite data. *Bollettino di Geofisica Teorica ed Applicata*, **XI**, 108–118.

Rapp, R. H., 1986. Global geopotential solutions. In Sünkel, H. (ed.), *Mathematical and Numerical Techniques in Physical Geodesy*. Berlin/Heidelberg: Springer. Lecture Notes in Earth Sciences, Vol. 7.

Rapp, R. H., and Cruz, J. Y., 1986a. *The representation of the Earth's gravitational potential in a spherical harmonic*

expansion to degree 250. Report 372, Department of Geodetic Science and Surveying, Ohio State University, Columbus.
Rapp, R. H., and Cruz, J. Y., 1986b. *Spherical harmonic expansions of the Earth's gravitational potential to degree 360 using 30′ mean anomalies*. Report 376, Department of Geodetic Science and Surveying, Ohio State University, Columbus.
Rapp, R. H., and Pavlis, N. K., 1990. The development and analysis of geopotential coefficient models to spherical harmonic degree 360. *Journal of Geophysical Research*, **95**(B13), 21885–21911.
Rapp, R. H., Wang, Y. M., Pavlis, N. K., 1991. *The Ohio State 1991 geopotential and sea surface topography harmonic coefficient models*. Report 410, Department of Geodetic Science and Surveying, Ohio State University, Columbus.
Schuh, W. D., 1996. *Tailored numerical solution strategies for the global determination of the Earth's gravity field*. Technical Report, Institute of Theoretical Geodesy, Technical University of Graz, Austria.
Sneeuw, N., 1994. Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective. *Geophysical Journal International*, **118**, 707–716.
Tscherning, C. C., 2001. Computation of spherical harmonic coefficients and their error estimates using least-squares collocation. *Journal of Geodesy*, **75**, 12–18.
Wang, Y. M., 1987. Numerical aspects of the solution of Molodensky's problem by analytical continuation. *Manuscripta Geodaetica*, **12**, 290–295.
Wang, Y. M., 1988. *Downward continuation of the free-air gravity anomalies to the ellipsoid using the gradient solution, Poisson's integral and terrain correction – numerical comparison and the computations*. Report 393, Department of Geodetic Science and Surveying, Ohio State University, Columbus.
Wenzel, H.-G., 1985. *Hochauflösende Kugelfunktionsmodelle für das Gravitationspotential der Erde. Wiss. Arb.* 137, Fachrichtung Vermess. der Univ. Hannover, Hannover, Federal Republic of Germany.
Wenzel, G., 1998. Ultra high degree geopotential models GPM98A, B and C to degree 1800. http://www.gik.uni-karlsruhe.de/~wenzel/gpm98abc/gpm98abc.htm.
Wenzel, G., 1999. Schwerefeldmodellierung durch ultrahochauflösende Kugelfunktionsmodelle. *Zeitschrift für Vermessungswesen*, **124**(5), 144–154.

Cross-references

Geoid
Gravity Method, Satellite
Gravity, Global Models
Gravity, Gradiometry
Magnetic, Global Anomaly Map
Numerical Methods, Boundary Element

STATISTICAL SEISMOLOGY

David A. Rhoades
GNS Science, Lower Hutt, New Zealand

Synonyms

Earthquake statistics; Statistical methods in seismology

Definition

Stochastic process. A process evolving in time governed partly by probability distributions.

Point process. Stochastic process of point events in time or space or both.
Renewal process. Point process in which the probability distribution of the time to the next event depends only on the time since the last event.
Stationary Poisson process. Point process in which the numbers of events in distinct time intervals of the same length are independent and have the same distribution.
Branching process. Process involving ancestors and offspring.

Introduction

Statistical seismology is an emerging field of research at the interface of statistical and physical modeling of earthquake occurrence. It is concerned with quantifying and understanding the distribution of earthquakes in time, magnitude, and location. It includes empirical analysis of earthquake catalogues, stochastic modeling of earthquake occurrence, estimation of the probability of earthquake occurrence and of earthquake-induced ground shaking, and testing the forecasting power of physical and statistical models of seismicity (Vere-Jones et al., 2005). It provides input to the management of seismic hazard and risk. Although efforts have so far focused mainly on the study of earthquake catalogues, there is scope to incorporate other geophysical measurements into the analyses.

Empirical relations

One of the classical empirical relations of statistical seismology is the Gutenberg and Richter (1944) law, according to which the distribution of earthquake magnitudes M within a large time-space volume has a probability density proportional to 10^{-bM}, where the parameter b is close to 1. Physical upper limits to earthquake size necessitate tapering of the density at high magnitudes (Kagan, 1999). For small space-time volumes and in certain tectonic settings, the b-value can differ markedly from 1. Temporal variation in b-value has been proposed as an earthquake precursor. Since the radiated seismic energy E_S of an earthquake is approximately proportional to $10^{1.5M}$, the Gutenberg–Richter law implies that the distribution of radiated seismic energy has density proportional to $E_S^{-(\beta+1)}$, where $\beta = 2/3b$. Thus, the Gutenberg–Richter law can be viewed as a power law. The existence of such power laws is often taken as evidence of fractality and self-organized criticality in the earthquake process (Bak and Tang, 1989).

Earthquakes tend to occur in clusters. A common type of cluster is the main shock – aftershock sequence. The Omori–Utsu law, first noted by Fusakichi Omori in 1894, describes how an aftershock sequence decays over time (Utsu et al., 1995). According to this relation, which is also a power law, the rate of aftershock occurrence is proportional to $(t + c)^{-p}$, where t is the elapsed time following the occurrence of the main shock, and c and p are adjustable parameters, typically with $c << 1$ day, and $p \approx 1$. In some large aftershock sequences, the relation

can be seen to hold for decades after the occurrence of the main shock.

The area occupied by an aftershock sequence is approximately 10^{M-4} km^2, where M is the main shock magnitude (Utsu, 1961). This is consistent with scaling relations derived for earthquake source dimensions – fault length, width, and displacement – all of which are approximately proportional to $10^{0.5M}$ (Wells and Coppersmith, 1994). However, sometimes major earthquakes appear to trigger aftershocks at long distances from the earthquake source. These can be accommodated by a power law in which the aftershock density is proportional to $(d^2 + r^2)^{-q}$, where r is the distance from the source, q is an adjustable parameter, typically about 1.5, and d is a function proportional to$10^{0.5M}$.

According to Båth's law (Båth, 1965), the largest aftershock is typically about 1.2 magnitude units smaller than the main shock. This indicates that the main shock and its aftershocks do not by themselves conform to the Gutenberg–Richter relation, the main shock being too large relative to the other earthquakes in the set. The existence of other types of earthquake clusters than main shock – aftershock sequences, such as swarms (which have no main shock) and multiplets (which may have several main shocks) is a confounding element in attempts to systematically examine Båth's relation.

Rules for defining and classifying earthquake clusters are necessarily somewhat arbitrary. Sometimes a stochastic model of earthquake clustering is used to calculate the probability that a given earthquake belongs to a cluster, and then the analysis considers many different possible groupings of the earthquakes into clusters (Zhuang et al., 2002).

Data quality is an ever-present issue affecting statistical analysis of earthquake catalogues. It is necessary to establish the quality of a catalogue at the outset, because an analysis can be seriously compromised by changes in the quality of the catalogue over time, such as temporal and spatial variation of the magnitude threshold of completeness. On the positive side, catalogue data are generally improving in both quality and quantity as better seismograph networks and earthquake location techniques are employed, and this is creating opportunities for more detailed analyses.

Precursors

Studies aimed at identifying precursors of large earthquake are faced with a dual challenge: Whereas large earthquakes are relatively rare, the number of degrees of freedom available when attempting to identify precursors is large. Therefore, a careful analysis is required before the conclusion is drawn that a proposed precursory phenomenon is real. A retrospective analysis can be used to define a possible precursory phenomenon, but prospective testing is necessary to confirm it.

Many major earthquakes are preceded in the long term by an increase in the rate of occurrence and magnitude of minor earthquakes, in an area not much larger than the major earthquake source. This is called the precursory scale increase (Evison and Rhoades, 2004). Sometimes, it consists of a sequence of swarms. The magnitude of the largest precursor is typically about one unit smaller than that of the main shock, and the precursor time T_P (the time from the onset of the increase to the main shock) and the area in which the precursors, major earthquake, and aftershocks all take place are both approximately proportional to $10^{0.5M}$. Earthquake precursors for which T_P is proportional to $10^{0.5M}$ are known as precursors of the first kind (Rikitake, 1982), and include anomalies in tilt, strain and land deformation, and changes in the relative velocities of seismic P and S waves. A distinguishing feature of the precursory scale increase is that the magnitude of the largest precursor(s) can be used to predict the time of occurrence, magnitude, and source area of the major earthquake.

Models of earthquakes as a critical phenomenon have been invoked to suggest that an accelerating moment release (AMR) should occur in an area much larger than the earthquake source in the approach of criticality, i.e., an accelerating occurrence of minor earthquakes leading up to a major earthquake. Another much-studied phenomenon is precursory seismic quiescence (PSQ) – a proposed reduction in the rate of occurrence of minor earthquakes shortly before a major event. There is apparent empirical support for both AMR and PSQ, but its statistical significance has been questioned. A problem in resolving such matters is that these phenomena, like many other proposed precursors, have only been defined anecdotally and never identified with an explicit stochastic model of earthquake occurrence.

Stochastic models of earthquake occurrence

Early examples of stochastic modeling of earthquake sequences were given by Vere-Jones (1970). Modeling of earthquake occurrence has benefited from the development of the theory of stochastic point processes (Daley and Vere-Jones, 2003).

The idea of successive episodes of stress accumulation and release, as suggested by the elastic rebound theory of Henry Fielding Reid in 1910, is the basis for most modeling of recurrence of major earthquakes on a fault or fault segment. The time sequence of earthquakes is often modeled as a renewal process, assuming a characteristic magnitude for events on the fault. The recurrence-time distribution is variously taken as exponential (for constant hazard), Weibull, lognormal, or Brownian passage-time. The latter three distributions all imply some degree of regularity in the time intervals between earthquakes, and physical arguments have been advanced in support of each. The limited data so far available on earthquake recurrence on fault segments do not support a clear preference for a particular distribution or afford a real opportunity to test how informative renewal models are in this context. It is important to account for data and parameter uncertainties when applying such models to the estimation of earthquake hazard.

The renewal process model does not allow for interactions between neighboring faults or for variation in the size of successive events. Models involving time and magnitude include the time-predictable and slip-predictable models, in which either the time to the next earthquake depends on the magnitude of the last earthquake or the magnitude of the next earthquake depends on the elapsed time. These models are sometimes applied to the earthquakes in a region. In a stress-release model of Zheng and Vere-Jones (1994), also applicable to the earthquakes in a region, the hazard rate steadily increases over time, but drops upon the occurrence of an earthquake by an amount that depends on its magnitude. A coupled version of this model, allowing for stress transfer between discrete regions, has also been developed. Further tests are needed of how well these models perform.

The epidemic-type aftershock (ETAS) model of Ogata (1989) is a nonstationary Poisson process model developed originally as a model of temporal clustering of earthquakes in a discrete region. In this model, each earthquake has its own aftershock sequence, which decays over time according to the Omori-Utsu law. The number of aftershocks is governed by a magnitude-dependent productivity function. Some earthquakes occur independently according to a stationary Poisson process, and the magnitudes of all earthquakes follow the Gutenberg–Richter relation. The ETAS model fits earthquake data much better than a stationary Poisson process, and better than a model in which only the larger events have aftershocks. It is a type of branching process model. A double branching process model has been found to better describe the time-variation of earthquake occurrence (Marzocchi and Lombardi, 2008).

Stochastic models involving location, as well as time and magnitude, allow for a more realistic representation of the earthquake process, including clustering and long-range interactions between earthquakes.

The ETAS model was extended to include the spatial variable (Ogata, 1998) by introducing a distribution for aftershock location, usually a power law of distance similar to that described above. In this form, it can be used to describe the space-time clustering of earthquakes and also as a diagnostic tool for physical changes affecting the earthquake process. As a short-term forecasting model, this model performs far better than smoothed seismicity models, which capture the space and magnitude distributions of past earthquakes but are time-invariant. It is successful at forecasting aftershocks, as well as the minor proportion of major earthquakes which are preceded by a foreshock sequence. For a given location, the rate density of earthquake occurrence in this model can fluctuate in a short time over several orders of magnitude. Changes in the parameters of the spatial ETAS model have been used as indirect indicators of stress-changes in regions of the crust.

The Every Earthquake a Precursor According to Scale (EEPAS) model (Rhoades and Evison, 2004) is a space-time-magnitude model designed for longer-term forecasting of the major earthquakes. In this model, every earthquake contributes to the future distribution of hazard in time, magnitude, and location on a scale determined by its magnitude, through the predictive scaling relations associated with the precursory scale increase phenomenon. Under this model, the rate density can fluctuate slowly in time over about 1.5 orders of magnitude. The EEPAS model outperforms smoothed seismicity models in well-catalogued inter-plate regions such as California, Japan, and New Zealand.

Testing of forecasting methods

Time-varying models of earthquake occurrence have not yet been widely adopted for practical purposes. The stationary Poisson model with characteristic earthquake magnitudes on faults and spatially distributed background seismicity conforming to the Gutenberg–Richter magnitude relation are the basis for most seismic hazard models in practical use. But an international effort by the Collaboratory for the Study of Earthquake Predictability (CSEP) to provide transparent, verifiable prospective tests of time-varying earthquake occurrence models is in progress. Regional earthquake forecast testing centers have been established in California, New Zealand, Europe and Japan, and others are planned. To be testable, a model must provide estimates of the expected number of earthquakes for future time windows within grid cells finely delimited by location and magnitude. Several different time steps are used; for example, 24 h, 3 months, and 5 years. An overview of models submitted for testing in California was given by Field (2007) and first results were presented by Schorlemmer et al. (2010).

The performance of a probabilistic model of earthquake occurrence on a target set of earthquakes in a catalogue independent of the one to which it was fitted, is conveniently measured by the increase in the log likelihood of the earthquake catalogue under the model compared with that under a reference model, such as a stationary Poisson model with spatial smoothing of the locations of past earthquakes.

The CSEP testing centers use additional likelihood-based tests, which aim to identify significant differences between the total number of earthquakes expected and the number observed, and the distributions in time, magnitude, and location of cell expectations and those of the targeted earthquakes (Schorlemmer et al., 2007).

Some earthquake forecasting methods, such as the M8 algorithm, are alarm-based rather than probabilistic. For alarm-based methods, the error diagram (Molchan, 1990), in which the proportion of time or space-time for which a certain level of alarm is exceeded is plotted against the proportion of unpredicted earthquakes, is a useful assessment tool.

Analysis of synthetic earthquake catalogues

Physics-based numerical models of earthquake occurrence (Ben Zion, 2008) can be used to generate synthetic earthquake catalogues incorporating accepted physical

elements such as stress transfer between faults by means of elastic, viscoelastic, and other interactions; frictional or other material instabilities acting on fault surfaces; increasing stress accumulation due to tectonic forces; and the detailed geometry of the fault surfaces. For synthetic catalogues, important physical variables, which cannot be measured in the real earth, such as the distribution of stress, can be tracked in detail. Statistical analysis of synthetic earthquake catalogues can reveal under what physical conditions a particular statistical model is likely to be appropriate and contribute to physical understanding of empirical relations derived from earthquake catalogues.

Conclusion

Statistical seismology is concerned with modeling the empirical observations of earthquake occurrence and testing models against observations while trying to throw light on the physics of earthquake generation. A rapid improvement in seismological databases has been a stimulus to recent developments in the field, and this state of affairs is likely to continue. Improved earthquake catalogues and other systematically collected data, such as earth deformation data derived from Global Positioning System (GPS) networks offer opportunities for enhanced physical understanding of the earthquake process and more informative stochastic models. At the same time, increased computing power allows for increased detail and complexity in physics-based modeling. The major challenge statistical seismology faces is to use the new data, together with physical insights from empirical studies and detailed physical modeling, to develop ever more informative stochastic models of earthquake occurrence that can be applied to mitigating earthquake hazard.

Bibliography

Bak, P., and Tang, C., 1989. Earthquakes as a self-organized critical phenomenon. *Journal of Geophysical Research*, **94**, 15635–15637.

Båth, M., 1965. Lateral inhomogeneities of the upper mantle. *Tectonophysics*, **2**, 483.

Ben-Zion, Y., 2008. Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes and different dynamic regimes. *Reviews of Geophysics*, **46**(4), RG4006.

Daley, D. J., and Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes*, 2nd edn. New York: Springer, Vol. 1.

Evison, F. F., and Rhoades, D. A., 2004. Demarcation and scaling of long-term seismogenesis. *Pure and Applied Geophysics*, **161**, 21–45.

Field, E. H., 2007. Overview of working group for the development of regional earthquake likelihood models (RELM). *Seismological Research Letters*, **78**, 7–16.

Gutenberg, B., and Richter, C. F., 1944. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, **34**, 185–188.

Kagan, Y. Y., 1999. Universality of the seismic moment-frequency relation. *Pure and Applied Geophysics*, **155**, 537–573.

Marzocchi, W., and Lombardi, A. M., 2008. A double branching model for earthquake occurrence. *Journal of Geophysical Research*, **113**, B08317, doi:10.1029/2007JB005472.

Molchan, G. M., 1990. Strategies in strong earthquake prediction. *Physics of the Earth and Planetary Interiors*, **61**, 84–98.

Ogata, Y., 1989. Statistical models for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics*, **169**, 159–174.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, **50**(2), 379–402.

Rhoades, D. A., and Evison, F. F., 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale. *Pure and Applied Geophysics*, **161**, 47–71.

Rikitake, T., 1982. *Earthquake Forecasting and Warning*. Center for Academic Publications, Tokyo/Reidel, Dordrecht, 402 p.

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., and Rhoades, D. A., 2007. Earthquake likelihood model testing. *Seismological Research Letters*, **78**, 17–29.

Schorlemmer, D., Zechar, J. D., Werner, M., Jordan, T. H., and the RELM Working Group, 2010. First results of the regional earthquake likelihood models experiment. *Pure and Applied Geophysics*, **167**, 859–876.

Utsu, T., 1961. A statistical study on the occurrence of aftershocks. *Geophysical Magazine*, **30**, 521–605.

Utsu, T., Ogata, Y., and Matsu'ura, R. S., 1995. The centenary of the Omori formula for a decay law of aftershock activity. *Journal of Physics of the Earth*, **43**, 1–33.

Vere-Jones, D., 1970. Stochastic models for earthquake occurrence. *Journal of the Royal Statistical Society*, **B 32**, 1–62.

Vere-Jones, D., Ben-Zion, Y., and Zuniga, R. (eds.), 2005. Statistical seismology, Pageoph Topical Volumes. *Pure and Applied Geophysics*, **162**(6/7), 1023–1396.

Wells, D. L., and Coppersmith, K. J., 1994. New empirical relations among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, **84**, 974–1002.

Zheng, X., and Vere-Jones, D., 1994. Further applications of stress release models to historical earthquake data. *Tectonophysics*, **229**, 101–121.

Zhuang, J., Ogata, Y., and Vere-Jones, D., 2002. Stochastic declustering of space-time earthquake occurrence. *Journal of the American Statistical Association*, **97**, 369–380.

Cross-references

Artificial Water Reservoir Triggered Earthquakes
Characteristic Earthquakes and Seismic Gaps
Earthquake Precursors and Prediction
Earthquake Prediction, M8 Algorithm
Earthquake, Magnitude
Earthquakes and Crustal Deformation
Earthquakes, Early and Strong Motion Warning
Earthquakes, Energy
Seismic Data Acquisition and Processing
Seismic Quiescence and Activation

SUBDUCTION ZONES

Geoffrey A. Abers
Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA

Definition

Subduction zones are convergent plate boundaries involving at least one oceanic plate. The oceanic plate descends

beneath the other into the deep mantle, forming a deep-sea trench, often a volcanic arc subparallel to it, an inclined Wadati–Benioff seismic zone, and other subparallel deforming belts.

Introduction

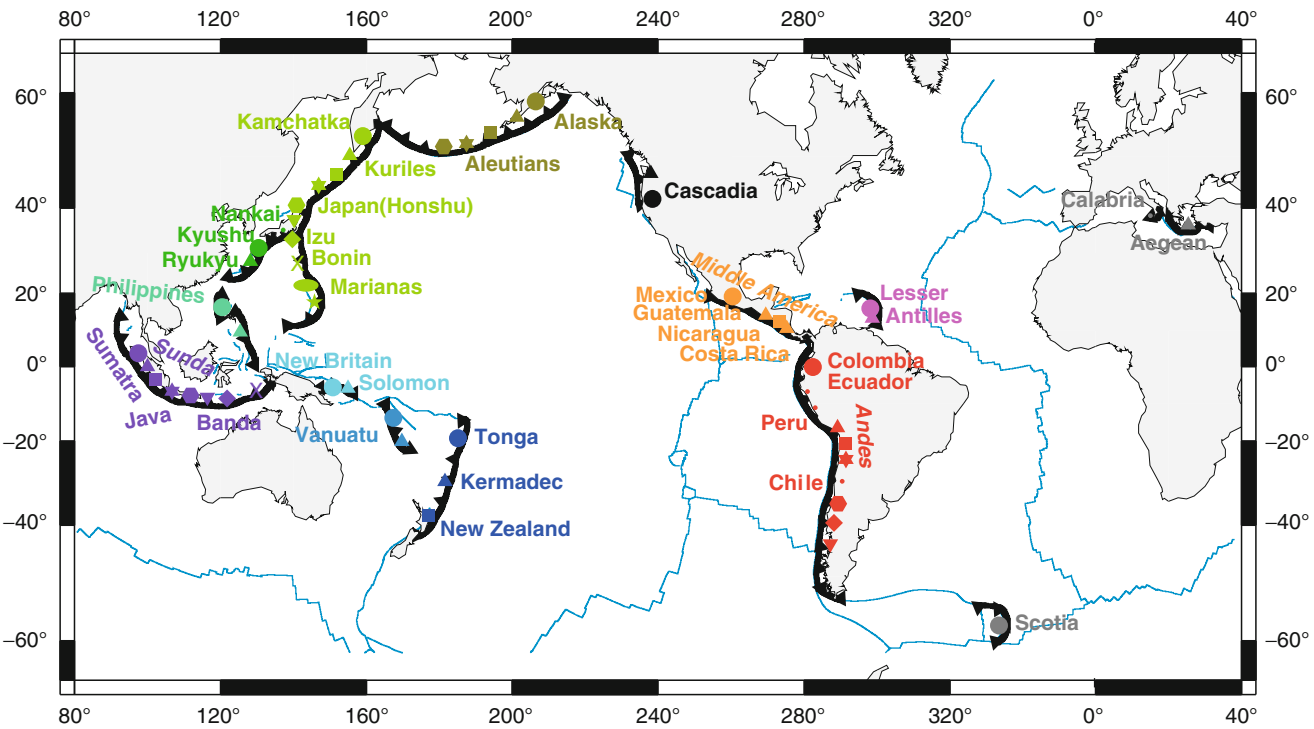
Subduction zones are a fundamental manifestation of planetary convection at the Earth's surface, and can be viewed as regulating two critical systems:

1. They control the primary material flux from the surface to the Earth's interior. Material at the seafloor subducts to depth, some of which emerges in volcanic arcs. These processes control the long-term budgets of volatiles (H_2O, CO_2, etc.), and the major accretion of material to continental crust.
2. Forces at subduction zones result in much of the planet's deformation, including the generation of the largest earthquakes and explosive volcanic eruptions over time scales of seconds, to the formation of the deepest seafloor and the growth of continents over millions of years. The negative buoyancy of subducting lithosphere probably represents the largest driver of global plate motions.

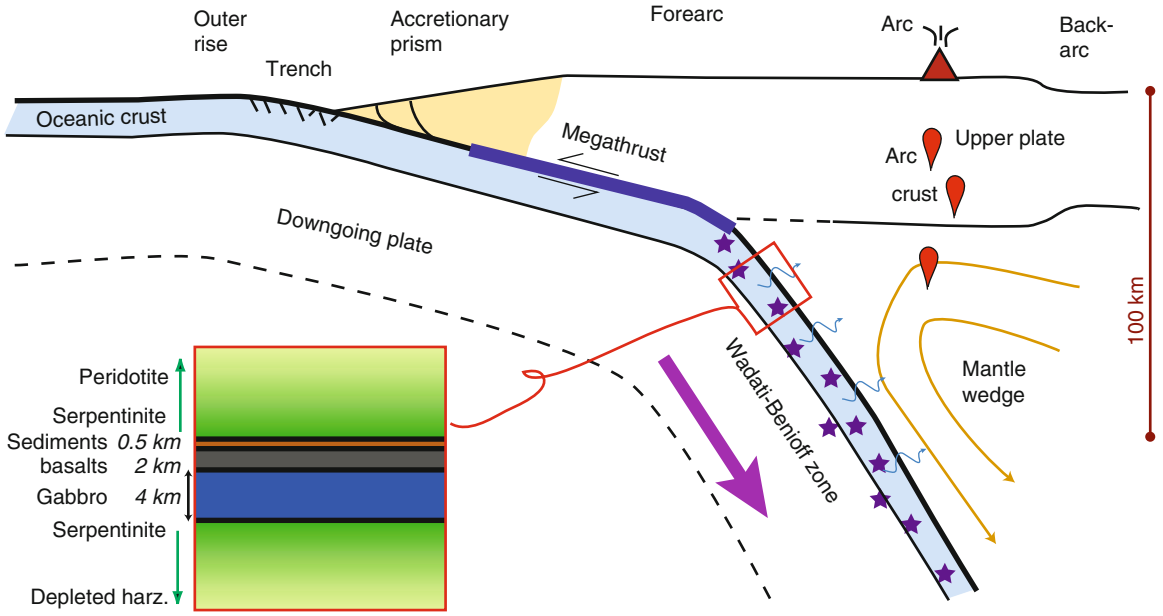
All of these processes ultimately result from the Earth releasing heat, and the deformation and thermal structure that results. At the largest scale, subduction regulates the mixing of material throughout the mantle of the lifetime of the planet, and the onset and termination of subduction represent profound geologic events. This article follows subduction zones from trench to depth, emphasizing the two themes of material flux and of large-scale deformation. It focuses on active systems from outer rise to sub-arc depths.

Morphology and nomenclature

The global subduction oceanic system is over 40,000 km long (Stern, 2002), largely but not entirely subducting oceanic plates of the Pacific basin (Figure 1). For the most part, any part of this subduction zone exhibits a common morphology (Figure 2): a downgoing, oceanic plate flexes into a deep-sea trench, beneath a submarine forearc that sometimes includes an accretionary prism. As the plate interface reaches 10–50 km depth, plate motion is accommodated on a shallow-dipping thrust zone, which in many places can generate great earthquakes. Landward of the thrust zone lies a volcanic arc, typically with a sharp front that overlies the region where the Wadati–Benioff zone seismicity reaches 75–180 km depth, with occasional volcanism further in the back-arc. The upper plate beneath and behind the arc varies between those that generate oceanic back-arc basins, exemplified by the Mariana-Izu-Bonin and Tonga systems, and those that undergo long-term convergence and compression, exemplified by the Andes, with a spectrum of behaviors in between (e.g., Lallemand et al., 2005). This characteristic morphology is a consequence of a regular, repeatable balance between forces that drive plate motions, buoyancy of



Subduction Zones, Figure 1 The global distribution of subduction zones. (Modified from Syracuse et al., 2010.)



Subduction Zones, Figure 2 Common features of subduction zones.

oceanic lithosphere, metamorphic devolatilization, and its rheological/magmatic consequences.

Outer rise, trench, and forearc

The shallow part of subduction systems, where temperatures are relatively cold, are controlled by the mechanics of elastic and brittle deformation. The trench occurs because the downgoing plate is flexed downward, and the shape of the plate seaward of the trench resembles closely that expected for flexure of a thin elastic surface (e.g., Watts and Talwani, 1974). The physics of elastic flexure also explains the outer rise – a region of slightly (100–500 m) elevated seafloor that lies seaward of the trench, typically 50–300 km distant (Parsons and Molnar, 1976). Strong bending leads to faulting in the outer rise, typically manifest as a series of normal faults with <1,000 m offset, seen on the seafloor and generating earthquakes (Chapple and Forsyth, 1979). These normal faults are found seaward of many if not most trenches, and are the most seaward indication of deformation within the subduction zone. They may play an important role in the material cycle as well, since the basins formed by them serve as traps for sediment that can be subducted to great depths (Hilde, 1983), and the faults themselves may act as conduits for water to reach into the mantle of the subducting plate, altering the mineralogy (e.g., Ranero et al., 2003). Such alteration of seafloor in the open ocean plays an important role in preconditioning the downgoing plate prior to subduction: the later metamorphic devolatilization of hydrous sediments and altered oceanic lithosphere releases fluids that lubricate the plate boundary, weaken the overlying material, and ultimately drive arc volcanism.

The sediment supply to the downgoing plate varies enormously between trenches near major continental river systems and those remote from such sources, leading to several variations in subduction zone behavior and characteristics. In the most extreme cases sediment supply overwhelms the rate at which material is subducted and accreted, so the bathymetric trench becomes a sediment-filled basin. Examples include the Lesser Antilles, Makran, and Cascadia subduction zones. A first-order distinction exists between subduction zones considered "accretionary," where a substantial amount of sediment has been offscraped to form an active accretionary prism, and "non-accretionary" subduction zones where all sediment is being subducted bypassing the forearc, and generally sediment supply is low (von Huene and Scholl, 1991). The accretionary prisms typically deform internally to form a critical-taper wedge, a geometry in which gravitational stresses created by the bathymetric slope are balanced by internal friction, similar to wedges in front of bulldozers or snowplows (Davis et al., 1983). Accreted material includes deep-sea sediment, but also fragments of bathymetric features such as seamounts on the incoming plate; subduction of relatively rigid seamounts can have a profound effect on the local morphology of the forearc, and may contribute to heterogeneity along the thrust zone at greater depth.

Thrust zone

As of mid-2010, the nine largest recorded earthquakes, and 22 of the largest 25, occurred on subduction thrust zones or "megathrusts," including all earthquakes with magnitudes of 9.0 and larger (U.S. Geol. Surv. Web site, http://earthquake.usgs.gov/). These great earthquakes also

generate most of the largest tsunamis, because they can displace large portions of the seafloor (see chapter on *Tsunami*). Great earthquakes occur here because the plate-boundary fault system can sustain large rupture areas, and earthquake size scales with rupture area. There are two reasons for the large rupture areas. First, at subduction zones cold material advects downward quickly compared with rates of heat conduction, as discussed below, so temperatures along the fault zone can remain below that of the brittle-ductile transition (or downdip limit of unstable sliding; Scholz, 1998) to greater depths than other environments, typically 30–50 km (Tichelaar and Ruff, 1993). Second, subduction zone thrusts dip at 5–25°, compared with steeper dips for faults in many other tectonic settings, so a greater fault area exists above the brittle-ductile transition. Both of these reasons allow for faults that can exceed 100 km in the downdip direction (e.g., for the Gulf of Alaska megathrust). Along strike, individual ruptures can extend for more than 1,000 km in some cases, most notably Chile (1960; Mw = 9.5) and Sumatra (2004; Mw = 9.2). Magnitudes here (Mw) are moment magnitudes (see chapter on *Earthquake, Magnitude*).

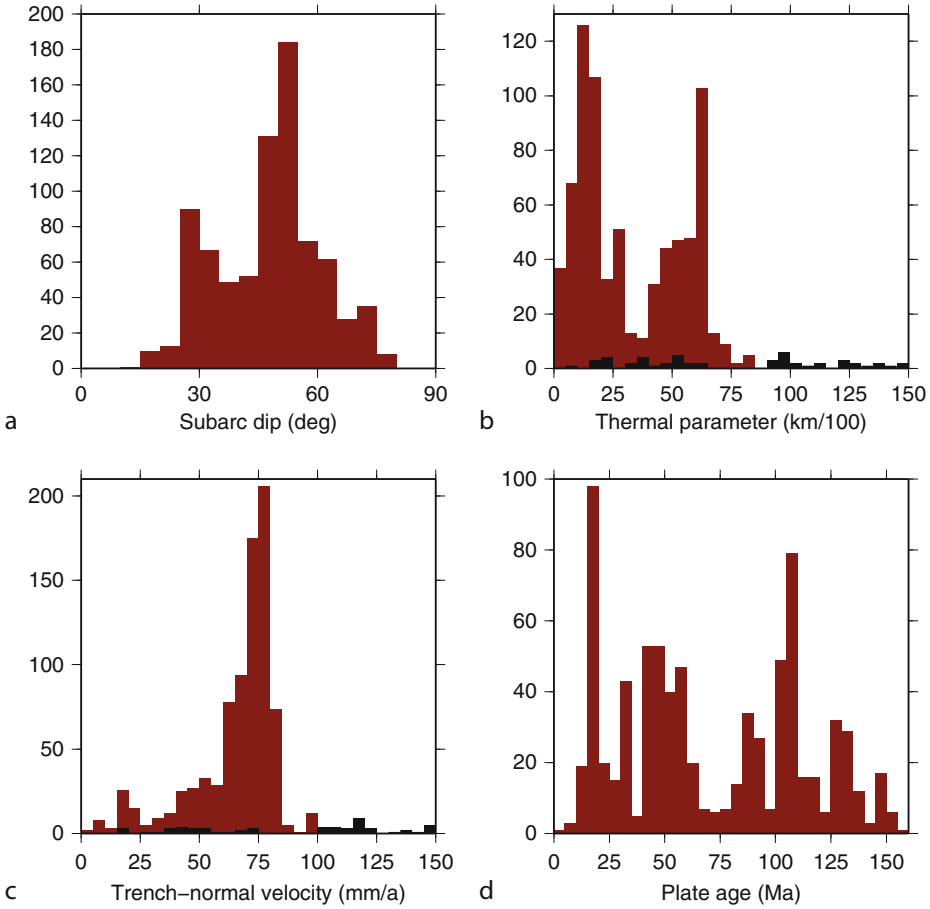
The fault area, and hence the ultimate size of earthquakes, is controlled by updip and downdip changes in material properties. At its updip end, the thrust zone undergoes a transition from aseismic slip on a decollement (subhorizontal plate-boundary fault) to stick–slip behavior or earthquake rupture, at depths of a few kilometers. At these depths the plate interface likely lies at temperatures of 75–150°C. The causes of this transition likely have to do with compaction, lithification, and low-temperature mineral phase transitions that change the frictional properties of this fault surface (Saffer and Marone, 2003). Similarly, the downdip limit can be in many places approximated by a threshold temperature, in this case near 350–500°C, with notable exceptions (e.g., Hyndman et al., 1997). Both transitions must be a consequence of change in frictional properties at this depth, with aseismic creep taking place at greater depths (e.g., Scholz, 1998). In between, the megathrust fault zone shows a variety of slip behaviors from purely creeping or "decoupled" to fully seismic or "coupled." In this usage, "coupling" refers to the ratio of fault slip inferred from summing the seismic moment of earthquakes over some time interval, to the total plate motion expected across that boundary over the same time interval. Coupled or "locked" thrust zones can be identified with geodesy, which shows the elastic strain accumulation building up between large earthquakes.

Recent discoveries of slow, silent creep events indicate that the downdip transition (and perhaps the updip one) may be much more complicated (e.g., Rogers and Dragert, 2003; Ito et al., 2007; Obara, 2002). Continuous geodetic observations have shown that patches of the thrust zone can slip over periods of days to months, in some cases accounting for a large fraction of the plate motion within the patch that slips (usually, adjacent to the locked zone). In some cases these events repeat quasi-periodically at time scales of months to years, although periodicity is not yet established for many regions of slow slip (Schwartz and Rokosky, 2007). Often, the slow slip is accompanied by a variety of exotic seismic phenomena, including nonvolcanic tremor, low-frequency earthquakes, very-low-frequency earthquakes, and creep (Ide et al., 2007). In most cases, the tremor seems to come from the plate boundary or just above it and shows motion consistent with thrusting on the plate boundary; these observations are subject of much current research (Gomberg, 2010). The physics of these quasi-periodic slip transients remain poorly understood; excess fluid pressure is suspected to play a role.

Kinematics

Subduction zones can be divided between those with little upper-plate deformation, with upper-plate compression, and exhibiting back-arc opening. The highest subduction rates, and indeed the highest relative plate motion rates anywhere, reach 240 mm/year at the northern end of the Tonga subduction system (Bevis et al., 1995). Major-plate motion between Australia and the Pacific Plate is only ~80 mm/year, the remainder being accommodated by spreading in the Lau back-arc system. Subduction can in principle be arbitrarily slow, although the lowest rates associated with an active arc probably occur in the Lesser Antilles (20–25 mm/year convergence; DeMets, 2001). A wide range of convergence rates lie in between these end-members with a sharp global peak near 70 mm/year such that half of all subduction zones have arc-normal convergence rates of 53–76 mm/year (Figure 3).

While spreading at mid-ocean ridges is orthogonal to their axes, at subduction zones convergence commonly trends obliquely to trench or arc (e.g., McCaffrey, 1992; Fitch, 1972). The obliquity, ϕ, is the angle between the normal to trench axis and the plate convergence direction. At low obliquities, convergence occurs in the direction of plate motion but commonly at higher obliquities strain is partitioned between a trench-normal component, represented by thrust faulting on the megathrust, and an along-strike component taken up by internal deformation. The along-strike component may be taken up within the forearc, as arc-parallel strike-slip systems, as sequences of strike-slip faults oblique to the trench axis, or by other means. At many subduction zones the obliquity changes along strike as both plate orientation and location relative to Euler pole change; for example, convergence is trench-normal ($\phi \sim 0$) in the Alaska Peninsula segment of the Aleutian subduction zone, $\phi \sim 45°$ from the trench in the central Aleutians, and plate motion is essentially parallel to the trench ($\phi \sim 90°$) in the far western Aleutians (DeMets et al., 1990). The latter case implies that material no longer descends into the mantle in the far western Aleutians, but translates along strike. A primary consequence of obliquity is that the trench-normal component of convergence can be significantly less than plate motion rates.



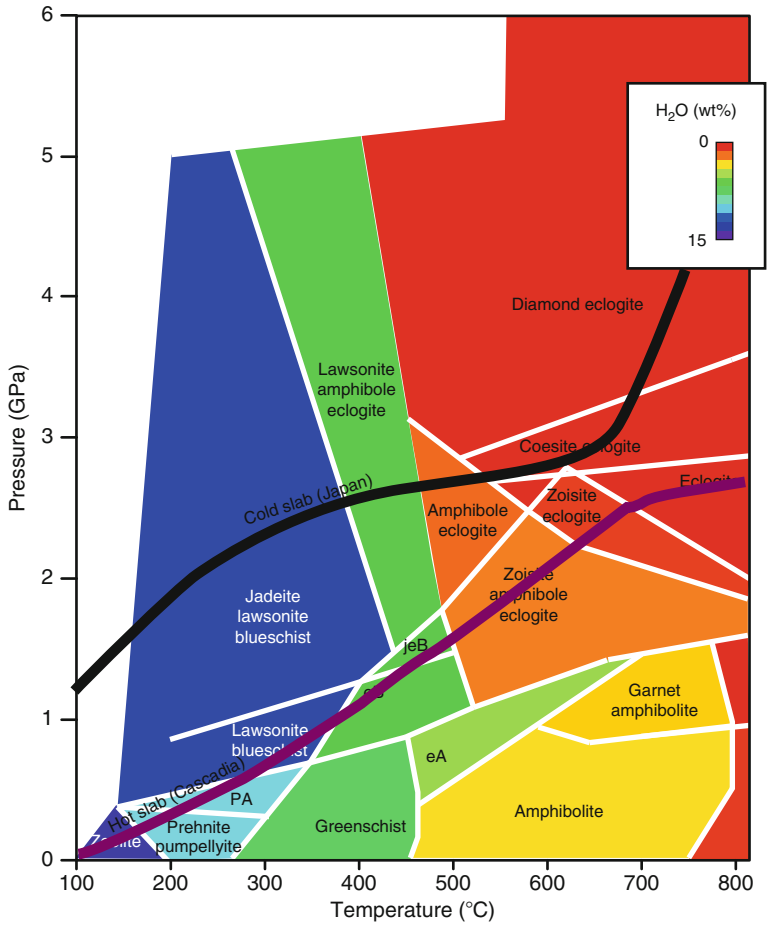
Subduction Zones, Figure 3 Histograms showing common parameters at subduction zones. Each is calculated at arc volcanoes or the nearest point on the plate boundary to them. (Based on data from Syracuse and Abers 2006.)

The kinematics of plates at asthenospheric depths may be more complex. While it is convenient to draw flow lines of subducting plates parallel to their dips, assuming steady-state slab geometry, it is not obvious that is the case. In many geodynamic models (e.g., Billen, 2008) slabs sink vertically faster than they would if they followed their trajectories, in other words, slabs roll back. Such rollback is probably required to accommodate back-arc opening in many places, although few if any constraints exist on these kinematics. Flow kinematics are complicated by inferences of large along-strike flow inferred from seismic anisotropy (e.g., Russo and Silver, 1994; Hoernle et al., 2008; see chapter on *Seismic Anisotropy*), not easily explained by simple models of wedge flow.

Thermal structure, plate buoyancy, mantle flow

At subduction zones, cold oceanic lithosphere advects downward relatively quickly compared with the rate of heat conduction, leading to some of the largest lateral temperature gradients in the planet. Old, fast-subducting slabs can retain temperatures $<600°C$ in their cores to depths exceeding 100 km, where ambient mantle temperatures are otherwise $\sim$1,400°C (e.g., Syracuse et al., 2010; Wada and Wang, 2009). The low temperatures lead to relatively large density differences, giving oceanic lithosphere significant negative buoyancy and probably supplying one of the major forces that drive plate motions (e.g., Forsyth and Uyeda, 1975). Because temperatures are low, mineral phases that hold significant H_2O are stable to much greater depths than elsewhere, making subduction zones major conveyors of volatiles to the deep earth. Still, volcanism occurs in nearly all oceanic subduction zones, indicating that parts of the system remain warm enough to produce melting. As a slab descends, it induces corner flow in the overlying mantle wedge, which draws warm material from the distal back-arc toward the slab corner (Figure 4). This advection of warm material ultimately heats the top of the slab, leading to very strong temperature gradients near the slab surface, and driving critical material-transport processes there.

Where the slab surface lies at depths less than 50 km, plate-boundary motion is mostly localized along the megathrust that defines the plate boundary, so the



Subduction Zones, Figure 4 Pressure-temperature trajectories for the top of slabs in hot (Cascadia) and cold (Japan) subduction zones, and metamorphic facies predicted for hydrated metabasalts, color coded by H_2O content. (Facies from Hacker et al. 2003; trajectories from Syracuse et al., 2010.)

upper-plate deformation accommodates little of the plate motion (at least the downdip component; oblique slip partitions strike-parallel motion into the upper plate). As a consequence, heat transport in this region can be approximated as the balance between conduction through an overlying, stationary upper plate, and downward advection of cold material along the thrust zone (Molnar and England, 1990). Shear heating along the thrust zone may elevate temperatures but its effect seems to be minor in the few places where abundant heat-flow measurements have been made (Cascadia, Japan, Tonga), equivalent to that generated by a fault with shear stress of a few tens of MPa at most (Hyndman and Wang, 1995; von Herzen et al., 2001). Thus, interplate thrust faults are relatively weak, perhaps due to high fluid pressures. One consequence is that metamorphism of subducted material follows a low-temperature high-pressure trajectory through this region, to blueschist facies (Figure 4).

At greater depth, the overlying material flows and transports heat from the back-arc mantle into the wedge. As this hot material is entrained downward by viscous drag, it rapidly heats the top of the downgoing plate, resulting in dehydration of the top of the downgoing plate. Such heating must happen under or trenchward of the volcanic arc, since many arcs erupt basaltic lavas with chemistry that requires temperatures in excess of 1,200°C beneath the arc (Kelemen et al., 2003a). The temperatures predicted for the top of the slab depend upon the assumed rheology of the mantle wedge, but for power-law flows typical of olivine dislocation creep, a narrow (<25–50 km wide) thermal boundary layer should rapidly develop at the top of the slab, leading to slab-surface temperatures of 700–1,000°C directly beneath the volcanic front (Kelemen et al., 2003a; van Keken et al., 2002; Syracuse et al., 2010). This heating appears to be sufficient to drive off most mineral-bound volatiles in the altered oceanic crust and sediment. More complicated modeling that allows secondary flow of buoyant slab-surface material into the wedge (e.g., Gerya and Yuen, 2003) might lead to lower temperatures, if such flows are sustainable.

Heat supplied by the mantle wedge continues to conduct into the slab, gradually warming its interior. Simple

thermal models (e.g., McKenzie, 1969; Molnar et al., 1979) predict that the maximum depth of isotherms should scale with $\Phi = AV\sin\delta$, where A is the age of the incoming lithosphere, V is its downdip velocity, and δ is the slab dip. The quantity Φ, sometimes called the "thermal parameter" (e.g., Kirby et al., 1996), describes in a simple way the overall thermal structure of slabs at great depth, and correlates well with the maximum depth of earthquakes over some depth ranges (Gorbatov and Kostoglodov, 1997), indicating a thermal control for the processes that cause deep earthquakes. Ultimately, the negative buoyancy of cold slabs and their ability to descend into the lower mantle must depend upon Φ as well. The negative thermal buoyancy of the slab counterbalances the positive buoyancy of subducting crust, since gabbro is much less dense than peridotite, such that the net buoyancy is a function of plate age and thickness of subducting crust (Cloos, 1993). For normal (5–8 km) thickness of oceanic crust, lithosphere attains negative buoyancy if more than 10–30 Ma old, while crust thicker than 15–25 km will always resist subduction for any observed age of oceanic crust.

The transition from localized slip along a plate boundary to distributed flow appears to correspond to where the slab reaches 70–90 km depth in many settings (Wada and Wang, 2009; Abers et al., 2006; Syracuse et al., 2010). Evidence for this transition comes from heat-flow observations at a small number of arcs (northern Japan and Cascadia), which show a step from very low forearc heat flow to relatively high sub-arc and back-arc heat flow, at this point. Also, seismic attenuation shows a sharp step at this depth, in a manner consistent with a cold "nose" of mantle material overlying the shallower portion of the slab, transitioning rapidly to hot mantle beneath the arc. The cause of this transition likely has to do with the maximum depth at which plate motion can be accommodated by a localized shear zone, and hence rheology of that shear zone; the location of arcs relative to the slab indicates a transition at temperatures about 550°C (Conder, 2005; Syracuse et al., 2010).

Fate of the downgoing plate: devolatilization and metamorphism

The subducting oceanic lithosphere can be divided into several layers (Figure 2, inset):

(a) Deep-sea sediment, of widely variable thickness and composition,
(b) Variably altered oceanic crust (AOC) – largely the upper, basaltic part where hydrothermal circulation has the largest effect,
(c) Less altered gabbroic lower crust, and
(d) Subducting mantle lithosphere some of which may be hydrated to varying degrees.

Although some of this package will be scraped off in accretionary margins, much probably survives to subduct to sub-arc depths. Arc geochemistry includes signatures of subducted sediment, with lava chemistry that varies systematically in some elements in ways that correlate with variations in sediment chemistry (Plank and Langmuir, 1998). This observation requires that most of the sediment, especially that deposited away from the trench, subduct to sub-arc depths. For example, lavas at several arcs include trace amounts of ^{10}Be, a radionuclide with ~1.6 Ma half-life formed in the atmosphere and deposited with the shallowest sediment (Tera et al., 1986), indicating that the shallowest deep-marine sediment subducts intact and the material reaches the earth's surface within a few Ma (a couple of half-lives). Finally, seismological evidence also supports the deep subduction of oceanic crust relatively intact, to generate images seen in receiver functions (e.g., Rondenay et al., 2008; Kawakatsu and Watada, 2007), high-frequency guided waves (Abers, 2000), and other mode conversions (e.g., Helffrich, 1996).

In its entirety, this package subducts on the order 2×10^9 Tg/Ma of water at the trench (Hacker, 2008), about half of which enters as pore water that is expelled by shallow compaction, and the remainder bound in minerals. Numerous drilling and sampling studies constrain well the H_2O content (and CO_2) of the subducting sediment and AOC, but lower sections are relatively less well known. Perhaps the largest uncertainty in these estimates lies in the water content of subducted mantle lithosphere, a potentially vast reservoir. Peridotite will hydrate to serpentinite and serpentine can hold ~14 wt% H_2O, giving this very thick layer potential to dominate the global water budget. However, it is not yet known how much H_2O reaches mantle depths; probably mantle hydration at mid-ocean ridges is minor because near-ridge mantle temperatures there should exceed that of serpentine stability (~600°C) but later faulting such as at fracture zones and the outer rise (see Thermal Structure, Plate Buoyancy, Mantle Flow) could be major conduits. Less is known about concentrations of CO_2 and other volatiles degassed by metamorphism. Depending upon serpentinite contributions, the volatile outputs of volcanic arcs either roughly match the inputs, implying near-complete volatile recycling, or are significantly less, implying a long-term flux of H_2O from the surface ocean to deep mantle.

Anhydrous oceanic crust can be expected to undergo a series of phase transitions from gabbro to eclogite. These result in a major density increase; gabbro has a density ~10% lower than that of peridotite at similar temperatures and pressures, so unmetamorphosed oceanic crust has buoyancy that resists subduction, while eclogite (dominated by garnet and pyroxenes) is significantly denser than peridotite. The density increase, in transforming from gabbro to eclogite, may contribute significantly to the negative buoyancy of slabs that drives plate tectonics. However, reaction rates for garnet growth (a major step in eclogite formation) within coarse-grained, anhydrous gabbro may be relatively slow, such that gabbro may persist metastably to considerable depth within subduction zones (Ahrens and Schubert, 1975). The presence of water should substantially enhance that reaction rate (e.g., Hacker, 1996).

If hydrated, basalt/gabbro (and metasediment) are expected to undergo a lengthy sequence of metamorphic

reactions that, generally, release H_2O (Figure 4). As a consequence, H_2O should be released over a wide range of depths from the thrust zone to sub-arc mantle, hydrating the overlying mantle wedge (Schmidt and Poli, 1998; Hacker, 2008). At typical to cold subduction conditions, oceanic crust can be expected to transition through greenschist and blueschist facies, capable of holding up to ~5 wt% H_2O, with major dehydration occurring in the transformation to eclogite during high-pressure heating (Hacker et al., 2003). In warmer subduction zones, defined as those with fairly young incoming plates and slow subduction rates (e.g., Cascadia, southern Japan), descending crust should follow a higher-temperature path perhaps through amphibolite facies, dehydrating as amphibole breaks down at ~2 GPa. Most thermal models with realistic mantle-wedge viscosity structure predict that subducting crust should dehydrate by depths of 100–200 km (e.g., Syracuse et al., 2010), at least in its shallow parts.

The mantle within the subducting plate potentially holds hydrous minerals as well, primarily serpentine but also chlorite, as might entrained material lying above the downgoing plate. These layers should also dehydrate, at temperatures of 600–700°C for serpentines at <200 km depth (e.g., Schmidt and Poli, 1998) with chlorite breaking down at temperatures about 200° higher. Serpentine dehydration temperatures are probably reached in all subduction zones on the slab surface and, hence, above it, although it is possible that a weakly hydrated zone of chlorite stability may persist past slab depths (e.g., Grove et al., 2009). Subducting mantle lithosphere in most subduction zones should be cold enough for serpentine to be stable past the arc (Syracuse et al., 2010), and may dehydrate gradually past that. Water persisting to depths past 200 km has the potential to remain within the mantle to considerable depth, in the "alphabet" phases of dense hydrous magnesium silicates, which have broad stability fields at near-slab conditions (Thompson, 1992). Water reaching the transition zone (410–660 km depth) may remain for a long time, as the dominant minerals ringwoodite and wadsleyite can retain wt% water as defects in their nominally anhydrous structure (Kohlstedt et al., 1996). This reservoir has potential to hold several times more water than the current surface ocean; thus, the ability of serpentinized mantle lithosphere to transport H_2O to great depths may be a dominant control on the Earth's long-term water budget.

As water is released, it ascends into the overlying mantle wedge, crust, back along the thrust zone, and through the forearc (Peacock, 1990), potentially hydrating it (Hyndman and Peacock, 2003). These pathways provide major controls on rheological properties in all these regions and on the behavior and composition of arc volcanoes. Also, it is often suggested that the dehydration process generates earthquakes at intermediate depths (70–150 km) in subduction zones (Green and Houston, 1995; Kirby et al., 1996; Hacker et al., 2003). These processes are discussed extensively in the chapter on seismicity-subduction zones.

Subduction channel

In many collisional mountain belts and accreted terranes, high-pressure, low-temperature rocks have reached the Earth's surface after having descended to great depths. In some "ultra-high pressure" (UHP) terranes, mineral assemblages or textures in these rocks record pressures in excess of 3–4 GPa (depths > 100 km), indicating that these rocks descended deeply in subduction zones before returning to the surface. In part to explain these observations, a "subduction channel" has been often proposed (Cloos and Shreve, 1988). The subduction of weak, hydrated metasediments and altered oceanic crust, and their subsequent shearing and interaction with metamorphic fluids, suggests the presence of a heterogeneous weak zone at the top of the downgoing plate. Initially, material in this channel will be dragged downward by coupling with the subducting plate, but it should be buoyant relative to overlying mantle, and in the right conditions a return flow could develop delivering high-grade metamorphic rocks to the surface. A weak zone is also indicated by the presence of a "cold nose" in many subduction zones in the forearc mantle wedge, a feature that probably requires a weak decoupling layer to depths of 70–90 km on the top of the downgoing plate (e.g., Wada and Wang, 2009). The presence of UHP rocks at the earth's surface indicates that the return-flow process or something similar probably happens in some instances, but it remains unclear if the subduction channel is a common feature of subduction zones.

Melting and the volcanic arc

Reviews elsewhere describe the geochemical evidence for the nature of melting beneath volcanic arcs and the generation of arc crust (e.g., Stern, 2002; Kelemen et al., 2003b), which briefly summarized here for their relevance to geophysical observations. Magmas emplaced and erupted at subduction zones span a wide range of compositions. Basalts are found at most subduction zones and indicate that primary melting occurs within the mantle beneath the volcanic arc, since basalts primarily form from melting of peridotites (e.g., Gill, 1981). These basalts often differ from those found at mid-ocean ridges or hot spots in several important chemical trends; in particular, many elements that are strongly abundant in subducting sediment and altered crust are found in abundance in many volcanic arc rocks in a manner that correlates with sediment chemistry (Plank and Langmuir, 1998). In some arcs there is strong evidence for andesites forming under high pressure, perhaps requiring melting of downgoing crust (e.g., Kelemen et al., 2003b).

Magmatic differentiation within the crust and mixing with preexisting arc basement result in a wide spectrum of volcanic output. Typically, seismic velocities within arcs show an increase with depth consistent with more silicic, felsic rocks at shallow depths (Vp = 6.0–6.4 km/s) underlain by more gabbroic rocks in the middle-to-lower crust (Vp = 7.0–7.4 km/s) (e.g., Suyehiro et al., 1996;

Shillington et al., 2004). Since they are petrologically fractionated from more basaltic parents, the andesitic or granitic rocks observed at mid-to-upper crustal levels should be balanced by more mafic silica-poor residuals, perhaps in the lower crust. However, such lower-crustal cumulates are not very abundant, as interpreted from seismic velocities, exposed arc sections, and within continents formed by arc accretion. In general, there is good reason to think that continents grew from accretion of arcs since they share many geochemical characteristics (Rudnick, 1995), but overall continents are significantly more silicic than primary arc basalts. This discrepancy probably requires a significant mass loss in the complementary, mafic cumulates that should form in arc lower crust. Where and how this mass loss occurs remains unclear; it may be that the lower crust in arcs occasionally founders or delaminates back into the mantle, since its density may be higher than that of peridotite, but many other possibilities exist. Geophysical observations of the lower crust and upper mantle have potential to place constraints on the formation and evolution of arc crust, and hence the origins of continental material (e.g., Kodaira et al., 2007; Tatsumi et al., 2008).

Primary magmas emerging from the mantle contain several wt% H_2O and elevated CO_2, several times higher than in mid-ocean ridge basalts (e.g., Hirschmann, 2006). Mantle melting beneath arcs often occurs in the presence of H_2O (Grove et al., 2006), leading to the frequent speculation that dehydration reactions within the subducting slab trigger melting in the overlying mantle wedge, and hence the location of dehydration within the slab controls the location of arc volcanism (e.g., Tatsumi, 1986). However, experiments on hydrous metabasalts show that dehydration occurs over a wide range of depths beginning much shallower than the sub-arc mantle wedge (e.g., Schmidt and Poli, 1998), so that H_2O should flux the mantle at many depths, not just under the volcanic arc. Also, primary, dry basalts have been found in some arcs, so H_2O is not always required for melting. It may be that the return flow (the incoming limb of wedge corner flow) sufficiently decompresses hot asthenosphere to trigger melting. The relative importance of wet (flux) melting and dry (decompression) melting within arcs remains poorly understood, as do the pathways that magmas take from their locus of formation to volcanic arcs. In particular, it is not obvious why the arc front is a narrow, well-defined feature while the conditions for melting should be more widely spread. The depth to slabs beneath the arc front averages 110 km but varies coherently by a factor of 2 from arc to arc, suggesting some other controlling process probably related to thermal structure (England et al., 2004; Syracuse and Abers, 2006).

Seismic imaging of the subduction zone

Although subduction zones exhibit a characteristic surface morphology of trench, forearc, volcanic arc, etc., the evidence that these associations represent subduction of deep material comes primarily from geophysical imaging of deep structure. Early observations of seismic wave propagation in Tonga showed that the Wadati–Benioff zone lies in a cold, low-attenuation structure unlike mantle elsewhere on the planet, and provided some of the first evidence that these were sites of descending cold lithosphere (Oliver and Isacks, 1967). Since then, methods of seismic tomography have improved greatly (see *Seismic Tomography*) as has the global data set of earthquake travel times, leading to a clear picture of P-wave high-velocity anomalies that show cold slabs descending through the upper mantle and in some cases into the lower mantle (e.g., van der Hilst et al., 1997). These data show slabs as seismic anomalies 4–10% faster than surrounding mantle (see compilation by Lay, 1997), consistent with thermal models that predict temperatures 400–800°C colder than surroundings. These images, based on global earthquake travel time data sets, show that in some cases slabs reach the mantle transition zone and flatten out, as if reaching a partial barrier to flow (e.g., Japan) while in other cases they pass through the transition zone delivering subducted material to the lower mantle.

Seismic recordings from networks and arrays in subduction zones provide strong constraints on the subduction zone structure of a scale relevant to slab metamorphism, arc petrogenesis, and wedge flow. For example, as dehydration occurs, densities increase as do seismic velocities; basalts are 10–15% slower than surrounding peridotites while eclogites have similar velocities. Furthermore, serpentinized peridotite can be tens of percent slower than peridotite particularly for shear waves, depending upon the extent of serpentinization (hydration). As a result, seismic imaging has potential to show where and how metamorphism within slabs occurs. For example, with migration of P-to-S scattered waves (see *Seismic, Receiver Function Technique*) subducting crust has been observed to ~45 km depth in Cascadia and depths exceeding 100 km in Alaska and Japan, consistent with expected depths of basalt dehydration and differences in thermal structure (Rondenay et al., 2008). Studies of earthquake travel times and signal attenuation also provide information on structure of the mantle wedge where melting initiates, and have shown a clear hot mantle wedge, with a variety of structures suggesting melt transport pathways. Probably the densest long-lived monitoring effort exists in Japan, although a variety of portable array experiments have illuminated subduction zones in parts of the Andes, Central America, Tonga, the Marianas, Alaska, Cascadia, Greece, and other locales.

One of the more complex suites of seismic observations comes from the anisotropic fabric within the mantle of subduction zones, which results in "splitting" of shear waves into orthogonally polarized fast and slow pulses (see *Seismic Anisotropy*). Mantle flow in the dislocation creep regime leads to lattice-preferred orientation of olivine, and olivine is moderately (~10%) anisotropic, leading to an association between seismic anisotropy and mantle flow. However, observations in many subduction zones lie counter to expectation (e.g., Russo and Silver, 1994).

Simple models of mantle flow in subduction zones predict "corner flow," in which mantle on either side of the descending plate is entrained via viscous drag (Figure 2). To fill the potential void left by the downward advection, mantle far from the subducting slab flows inward toward it, eventually turning and flowing downward. While details of this flow depend upon the viscosity structure and its sensitivity to temperature, stress, water, and other variables, the overall pattern of flow is one in which the maximum strain or stretch should be perpendicular to trench and arc, in the direction of flow. In many subduction zones, anisotropic fabric does not show this pattern, and instead the fast direction for shear wave propagation is more often than not parallel to the trench, at least below and behind the volcanic arc (Fischer et al., 1998). More complicated fabrics are also observed, with sharp 90° changes in fabric between sub-arc mantle and the shallower part of the forearc (e.g., in North Japan and Alaska), and the pattern varies between different subduction zones. Explanations for these unexpected patterns remain a matter of much debate, and include strong along-strike flow, secondary flows within the wedge, influence of melt on fabric, activation of unusual olivine slip systems at some subduction zone conditions, abundance of hydrous, foliated minerals in the forearc, extensive fabric within the downgoing slab, and unexpectedly strong upper-plate fabrics.

Subduction zone hazards

The forces present in subduction zones also produce some of the most violent natural hazards known. The interplate thrust zone generates all of the largest earthquakes, which produce strong shaking but also submarine landslides, meter-scale uplift and subsidence, and most of the largest tsunamis. The 2004 Sumatra earthquake represents the primary example of the last 50 years; a magnitude 9.2 earthquake ruptured over 1,000 km of plate boundary, producing an ocean-crossing tsunami that directly resulted in over 150,000 deaths and huge devastation to property. These earthquakes are devastating because subduction zone thrust faults are very large, because they are submarine but can break the sea surface, and because they occur in places with often oversteepened bathymetry. Subduction zones are also site to many of the most violent volcanic eruptions. Large volatile (especially water) contents of subduction zone magmas, compared with most other settings, lead to magmas rich in gas that exsolve at shallow depths. Magmatic differentiation leads to relatively viscous, felsic magmas that rise slowly and do not degas easily. The combination leads to explosive and sometimes very large eruptions; in the last half century, Mt St Helens (Cascadia), Pinatubo (Philippines), and Montserrat (Lesser Antilles) all produced violent and well-documented eruptions in subduction zones, among many others. Worldwide, about 500 volcanoes have known historic eruptions, roughly two-thirds in subduction zones.

Summary

Subduction zones represent the primary return of material from the earth's surface to its interior. In the process of doing so, great earthquakes occur, violent volcanoes erupt, and the largest lateral temperature gradients in the Earth's interior are generated. At shallow depths, elastic and brittle processes control deformation including most of the planet's large earthquakes, while at greater depths the subduction system deforms by large-scale ductile flow. Along with oceanic crust, sediment and volatiles enter the trench, lubricating the plate boundary at shallow depths, and then are metamorphosed where slabs descend below hot mantle, releasing H_2O and other volatiles into the overlying mantle wedge. While some of this material may mechanically ascend back to the surface, in the form of high-pressure metamorphic rocks, much feeds explosive volcanism in the volcanic arc or descends into the deep mantle. Melting at sub-arc depths is controlled by a combination of advection via flow of hot mantle from the back-arc, and fluxing by slab-derived fluids, a combination of which produces a sharp volcanic front. Magmas ascend to form arc crust and differentiate, some fraction of which persists to form the basis of continental crust. This transfer of energy and material leads to a rich set of phenomena that play a major role in creating the geologic record.

Bibliography

Abers, G. A., 2000. Hydrated subducted crust at 100–250 km depth. *Earth and Planetary Science Letters*, **176**, 323–330.

Abers, G. A., van Keken, P. E., Kneller, E. A., Ferris, A., and Stachnik, J. C., 2006. The thermal structure of subduction zones constrained by seismic imaging: implications for slab dehydration and wedge flow. *Earth and Planetary Science Letters*, **241**, 387–397.

Ahrens, T. J., and Schubert, G., 1975. Gabbro-eclogite reaction rate and its geophysical significance. *Reviews of Geophysics and Space Physics*, **13**, 383–400.

Bevis, M., et al., 1995. Geodetic observations of very rapid convergence and back-arc extension at the Tonga arc. *Nature*, **374**, 249–251.

Billen, M. I., 2008. Modeling the dynamics of subducting slabs. *Annual Review of Earth and Planetary Sciences*, **36**, 325–356.

Chapple, W. M., and Forsyth, D. W., 1979. Earthquakes and bending of plates at trenches. *Journal of Geophysical Research*, **84**, 6729–6749.

Cloos, M., 1993. Lithospheric buoyancy and collisional orogenesis; subduction of oceanic plateaus, continental margins, island arcs, spreading ridges, and seamounts. *Geological Society of America Bulletin*, **105**, 715–737.

Cloos, M., and Shreve, R. L., 1988. Subduction-channel model of prism accretion, melange formation, sediment subduction, and subduction erosion at convergent plate margins: 1. Background and description. *Pure and Applied Geophysics*, **128**, 455–500.

Conder, J. A., 2005. A case for hot slab surface temperatures in numerical viscous flow models of subduction zones with an improved fault zone parameterization. *Physics of the Earth and Planetary Interiors*, **149**, 155–164.

Davis, D., Suppe, J., and Dahlen, F. A., 1983. Mechanics of fold-and-thrust belts and accretionary wedges. *Journal of Geophysical Research*, **88**, 1153–1172.

DeMets, C., 2001. A new estimate for present-day Cocos-Caribbean plate motion; implications for slip along the central American volcanic arc. *Geophysical Research Letters*, **28**, 4043–4046.
DeMets, C., Gordon, R., Argus, D., and Stein, S., 1990. Current plate motions. *Geophysical Journal International*, **101**, 425–478.
England, P., Engdahl, E. R., and Thatcher, W., 2004. Systematic variations in the depths of slabs beneath arc volcanoes. *Geophysical Journal International*, **156**, 377–408.
Fischer, K. M., Fouch, M. J., Wiens, D. A., and Boettcher, M. S., 1998. Anisotropy and flow in Pacific subduction zone back-arcs. *Pure and Applied Geophysics*, **151**, 463–475.
Fitch, T. J., 1972. Plate convergence, transcurrent faults and internal deformation adjacent to southeast Asia and the western Pacific. *Journal of Geophysical Research*, **77**, 4432–4460.
Forsyth, D. W., and Uyeda, S., 1975. Relative importance of driving forces of plate motion. *Geophysical Journal of the Royal Astronomical Society*, **43**, 163–200.
Gerya, T. V., and Yuen, D. A., 2003. Rayleigh-Taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones. *Earth and Planetary Science Letters*, **212**, 47–62.
Gill, J. B., 1981. *Orogenic andesites and plate tectonics*. New York: Springer, p. 390.
Gomberg, J., 2010. Slow-slip phenomena in Cascadia from 2007 and beyond: a review. *Geological Society of America Bulletin*, **122**, 963–978.
Gorbatov, A., and Kostoglodov, V., 1997. Maximum depth of seismicity and thermal parameter of the subducting slab: general empirical relation and its application. *Tectonophysics*, **277**, 165–187.
Green, H. W., and Houston, H., 1995. The mechanics of deep earthquakes. *Annual Review of Earth and Planetary Sciences*, **23**, 169–213.
Grove, T. L., Chatterjee, N., Parman, S. W., and Médard, E., 2006. The influence of H_2O on mantle wedge melting. *Earth and Planetary Science Letters*, **249**, 74–89.
Grove, T. L., Till, C. B., Lev, E., Chatterjee, N., and Medard, E., 2009. Kinematic variables and water transport control the formation and location of arc volcanoes. *Nature*, **459**, 694–697.
Hacker, B. R., 1996. Eclogite formation and the rheology, buoyancy, seismicity, and H_2O content of oceanic crust. In Bebout, G. E., Scholl, D., Kirby, S., and Platt, J. P. (eds.), *Subduction: Top to Bottom*. Washington, DC: American Geophysical Union. AGU Monograph Series, Vol. 96, pp. 337–346.
Hacker, B. R., 2008. H_2O subduction beyond arcs. *Geochemistry Geophysics Geosystems* **9**: art. No. Q03001.
Hacker, B. R., Abers, G. A., and Peacock S. M., 2003. Subduction factory 1: theoretical mineralogy, density, seismic wavespeeds, and H_2O content. Journal of Geophysical Research, **108**: art. No. 2029.
Helffrich, G., 1996. Subducted lithospheric slab velocity structure: observations and mineralogical inferences. In Bebout, G. E., Scholl, D., Kirby, S., and Platt, J. P. (eds.), *Subduction: Top to Bottom*. Washington, DC: American Geophysical Union. AGU Monograph Series, Vol. 96, pp. 215–222.
Hilde, T. W. C., 1983. Sediment subduction versus accretion around the Pacific: convergence and subduction. *Tectonophysics*, **99**, 381–397.
Hirschmann, M. M., 2006. Water, melting, and the deep earth H_2O cycle. *Annual Review of Earth and Planetary Sciences*, **34**, 629–653.
Hoernle, K., Abt, D. L., Fischer, K. M., Nichols, H., Hauff, F., Abers, G. A., van den Bogaard, P., Heydolph, K., Alvarado, G., Protti, M., and Strauch, W., 2008. Arc-parallel flow in the mantle wedge beneath Costa Rica and Nicaragua. *Nature*, **451**, 1094–1098.
Hyndman, R. D., and Peacock, S. M., 2003. Serpentinization of the forearc mantle. *Earth and Planetary Science Letters*, **212**, 417–432.
Hyndman, R. D., and Wang, K., 1995. The rupture zone of Cascadia great earthquakes from current deformation and the thermal regime. *Journal of Geophysical Research*, **100**, 22133–22154.
Hyndman, R. D., Yamano, M., and Oleskevich, D. A., 1997. The seismogenic zone of subduction thrust faults. *Island Arc*, **6**, 244–260.
Ide, S., Beroza, G., Shelly, D. R., and Uchide, T., 2007. A scaling law for slow earthquakes. *Nature*, **447**, 76–79.
Ito, Y., Obara, K., Shiomi, K., Sekine, S., and Hirose, H., 2007. Slow earthquakes coincident with episodic tremors and slow slip events. *Science*, **315**, 503–506.
Kawakatsu, H., and Watada, S., 2007. Seismic evidence for deep-water transportation in the mantle. *Science*, **316**, 1468–1471.
Kelemen, P. B., Rilling, J. L., Parmentier, E. M., Mehl, L., and Hacker, B. R., 2003a. Thermal structure due to solid-state flow in the mantle wedge beneath arcs. In Eiler, J. M. (ed.), *Inside the Subduction Factory*. Washington, DC: AGU. Geophysical Monograph, Vol. 138, pp. 293–311.
Kelemen, P. B., Hanghøj, K., and Greene, A. R., 2003b. One view of the geochemistry of subduction related magmatic arcs, with emphasis on primitive andesite and lower crust. In Rudnick, R. L. (ed.), *The Crust*. Oxford: Elsevier. Treatise on Geochemistry, Vol. 3.
Kirby, S., Engdahl, E. R., and Denlinger, R., 1996. Intermediate-depth intraslab earthquakes and arc volcanism as physical expressions of crustal and uppermost mantle metamorphism in subducting slabs. In Bebout, G. E., Scholl, D., and Kirby, S. (eds.), *Subduction: Top to Bottom*. Washington, DC: Amererican Geophysical Union. Geophysical Monograph, Vol. 96, pp. 195–214.
Kodaira, S., Sato, T., Takahashi, N., Miura, S., Tamura, Y., Tatsumi, Y., and Kaneda, Y., 2007. New seismological constraints on growth of continental crust in the Izu-Bonin intra-oceanic arc. *Geology*, **35**, 1031–1034.
Kohlstedt, D. L., Keppler, H., and Rubie, D. C., 1996. Solubility of water in the alpha, beta and gamma phases of (Mg, Fe)(2)SiO_4. *Contributions to Mineralogy and Petrology*, **123**, 345–357.
Lallemand, S., Hueuret, A., and Boutelier, D., 2005. On the relationships between slab dip, back-arc stress, upper plate absolute motion, and crustal nature in subduction zones. *Geochemistry Geophysics Geosystems*, **6**: art. no. Q090006.
Lay, T., 1997. *Structure and Fate of Subducting Slabs*. New York: Academic, p. 185.
McKenzie, D. P., 1969. Speculations on the consequences and causes of plate motions. *Geophysical Journal of the Royal Astronomical Society*, **18**, 1–32.
McCaffrey, R., 1992. Oblique plate convergence, slip vectors, and forearc deformation. *Journal of Geophysical Research*, **97**, 8905–8915.
Molnar, P., and England, P. C., 1990. Temperatures, heat flux, and frictional stress near major thrust faults. *Journal of Geophysical Research*, **95**, 4833–4856.
Molnar, P., Freedman, D., and Shih, J. S. F., 1979. Lengths of intermediate and deep seismic zones and temperatures in downgoing slabs of lithosphere. *Geophysical Journal of the Royal Astronomical Society*, **56**, 41–54.
Obara, K., 2002. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science*, **296**, 1679–1681.
Oliver, J., and Isacks, B., 1967. Deep earthquake zones, anomalous structures in the upper mantle, and the lithosphere. *Journal of Geophysical Research*, **72**, 4259–4275.
Parsons, B., and Molnar, P., 1976. The origin of outer topographic rises associated with trenches. *Geophysical Journal of the Royal Astronomical Society*, **45**, 707–712.
Peacock, S. M., 1990. Fluid processes in subduction zones. *Science*, **248**, 329–337.
Plank, T., and Langmuir, C. H., 1998. The chemical composition of subducting sediment and its consequences for the crust and mantle. *Chemical Geology*, **145**, 325–394.

Ranero, C., Morgan, J., McIntosh, K., and Reichert, C., 2003. Bending-related faulting and mantle serpentinization at the middle America trench. *Nature*, **425**, 367–373.
Rogers, G., and Dragert, H., 2003. Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip. *Science*, **300**, 1942–1943.
Rondenay, S., Abers, G. A., and van Keken, P. E., 2008. Seismic imaging of subduction zone metamorphism. *Geology*, **36**, 275–278.
Rudnick, R. L., 1995. Making continental crust. *Nature*, **378**, 571–578.
Rudnick, R. L., and Fountain, D. M., 1995. Nature and composition of the continental crust; a lower crustal perspective. *Reviews of Geophysics*, **33**, 267–309.
Russo, R. M., and Silver, P. G., 1994. Trench-parallel flow beneath the Nazca plate from seismic anisotropy. *Science*, **263**, 1105–1111.
Saffer, D. M., and Marone, C., 2003. Comparison of smectite- and illite-rich gouge frictional properties: application to the updip limit of the seismogenic zone along subduction megathrusts. *Earth and Planetary Science Letters*, **215**, 219–235.
Schmidt, M. W., and Poli, S., 1998. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth and Planetary Science Letters*, **163**, 361–379.
Scholz, C. H., 1998. Earthquakes and friction laws. *Nature*, **391**, 37–42.
Schwartz, S. Y., and Rokosky, J. M., 2007. Slow slip events and seismic tremor at circum-pacific subduction zones. *Reviews of Geophysics*, **45**: art. no. RG3004.
Shillington, D., Van Avendonk, H. J. A, Holbrook, W. S., Kelemen, P. B., Hornbach, M. J., 2004. Composition and structure of the central Aleutian island arc from arc-parallel wide-angle seismic data. *Geochemistry Geophysics Geosystems*, **5**: Art no. Q10006.
Stern, R.J., 2002. Subduction zones. *Reviews of Geophysics*, **40**: art. no. 1012.
Suyehiro, K., Takahashi, N., Ariie, Y., Yokoi, Y., Hino, R., Shinohara, M., Kanazawa, T., Hirata, N., Tokuyama, H., and Taira, A., 1996. Continental crust, crustal underplating, and low-Q upper mantle beneath an oceanic island arc. *Science*, **272**, 390–392.
Syracuse, E. M., and Abers, G. A., 2006. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochemistry Geophysics Geosystems*, **7**: art. no. Q05017.
Syracuse, E. M., van Keken P. E., and Abers G. A., 2010. The global range of subduction zone thermal models. *Physics of the Earth and Planetary Interiors*, **183**, 73–90.
Tatsumi, Y., 1986. Formation of the volcanic front in subduction zones. *Geophysical Research Letters*, **13**, 717–720.
Tatsumi, Y., Shukuno H., Tani K., Takahashi N., Kodaira S., and Kogiso, T., 2008. Structure and growth off the Izu-Bonin-Mariana arc crust: 2. Role of crust-mantle transformation and the transparent Moho in arc crust evolution. *Journal of Geophysical Research.***113**: art. no. B02203.
Tera, F., Brown, L., Morris, J., Sacks, I. S., Klein, J., and Middleton, R., 1986. Sediment incorporation in island-arc magmas: inferences from ^{10}Be. *Geochimica et Cosmochimica Acta*, **50**, 535–550.
Thompson, A. B., 1992. Water in the Earth's upper mantle. *Nature*, **358**, 295–302.
Tichelaar, B. W., and Ruff, L. J., 1993. Depth of seismic coupling along subduction zones. *Journal of Geophysical Research*, **98**, 2017–2037.
Van der Hilst, R. D., Widyantoro, S., and Engdahl, E. R., 1997. Evidence for deep mantle circulation from global tomography. *Nature*, **386**, 578–584.
van Keken, P. E., Kiefer, B., and Peacock, S. M., 2002. High-resolution models of subduction zones: implications for mineral dehydration reactions and the transport of water into the deep mantle. *Geochemistry Geophysics Geosystems*, **3**: art. no. 1056.
Von Herzen, R., Ruppel, C., Molnar, P., Nettles, M., Nagihara, S., and Ekstrom, G., 2001. A constraint on the shear stress at the Pacific-Australia plate boundary from heat flow and seismicity at the Kermadec forearc. *Journal of Geophysical Research*, **106**, 6817–6833.
von Huene, R., and Scholl, D. W., 1991. Observations at convergent margins concerning sediment subduction, subduction erosion, and the growth of continental crust. *Reviews of Geophysics*, **29**, 279–316.
Wada, I., and Wang, K., 2009. Common depth of slab-mantle decoupling: reconciling diversity and uniformity of subduction zones. *Geochemistry Geophysics Geosystems*, **10**: Q10009
Watts, A. B., and Talwani, M., 1974. Gravity anomalies seaward of deep-sea trenches and their tectonic implications. *Geophysical Journal of the Royal Astronomical Society*, **36**, 57–90.

Cross-references

Earthquake, Magnitude
Seismic Anisotropy
Seismic Tomography
Seismic, Receiver Function Technique
Tsunami

SURFACE WAVES

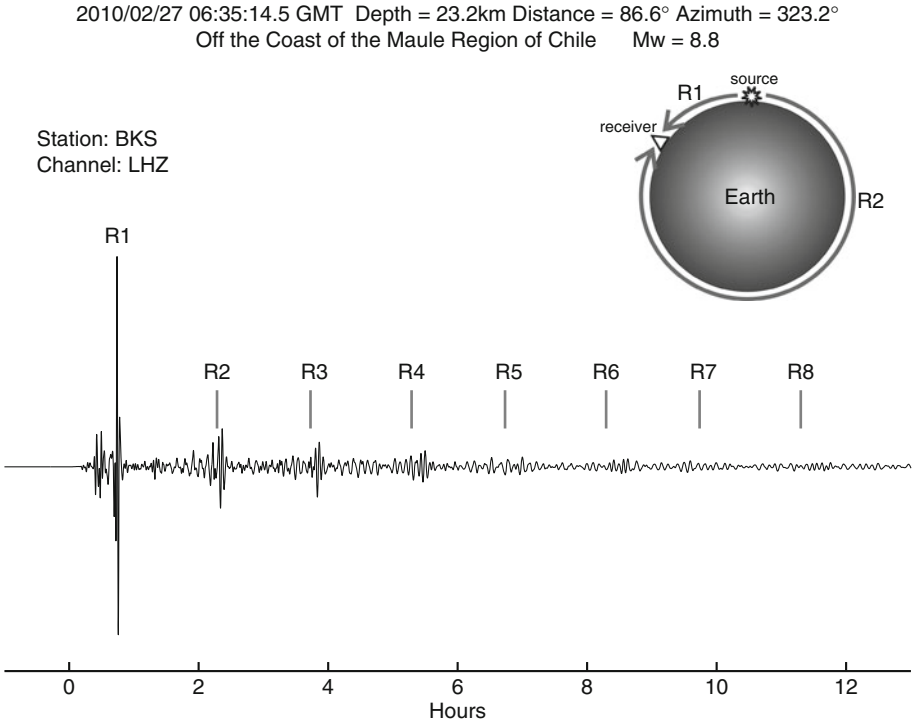
Barbara Romanowicz
Berkeley Seismological Laboratory, Berkeley, CA, USA

Definition

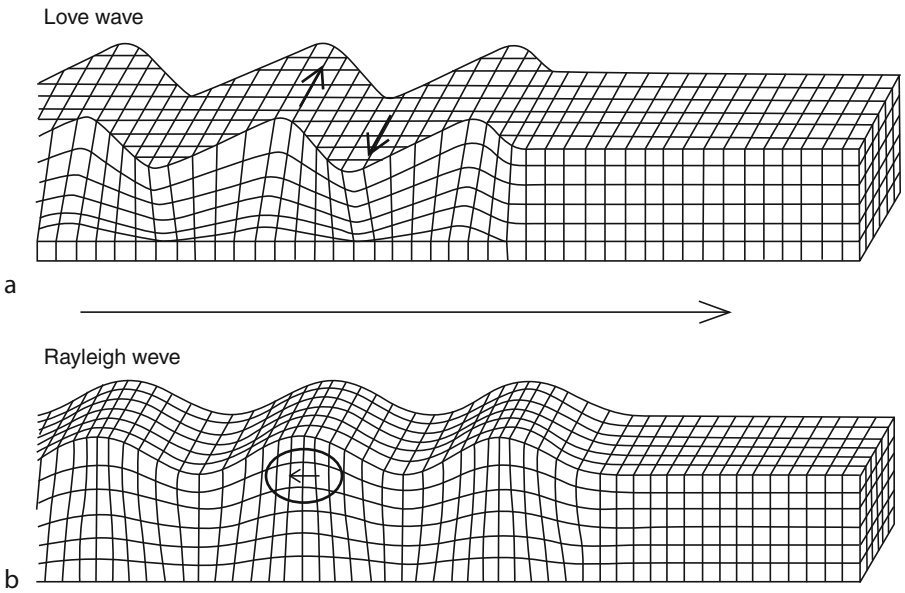
Surface waves are elastic waves which propagate along the surface of the earth and whose energy decays exponentially with depth.

Surface waves contain most of the long period energy (periods greater than 20 s) generated by earthquakes and recorded at teleseismic distances. Most prominent on records of moderate ($M \geq 5.5$) earthquakes are the fundamental mode wave trains that have propagated along the direct great circle path between the epicenter and the station. The dispersive and attenuative properties of these wave trains have been used extensively, since the 1950s, to infer crust and upper mantle structure at the regional scale. For earthquakes of magnitude 7 or larger, successive, earth-circling surface wave trains can be followed for many hours (Figure 1) and are either analyzed individually or, at the longest periods ($T > 250$ s) they are combined over time lengths of tens of hours or days to produce a spectrum of Earth's free oscillations (see *Free Oscillations of the Earth*).

In an isotropic medium, there are two types of surface waves which propagate independently (Figure 2). Rayleigh waves, polarized in the vertical plane containing the source and the receiver (P-SV energy), have a retrograde elliptical particle motion. These waves have generally good signal-to-noise ratio on the vertical component records and have been studied the most. On the other hand, the analysis of Love waves, polarized



Surface Waves, Figure 1 Example of vertical component record showing earth-circling Rayleigh wave trains (marked R1...R8) following the M8.8 Maule Chile earthquake of February 27, 2010. This record is from station BKS of the Berkeley Digital Seismic Network (BDSN). The successive wavepackets are well separated from each other at this epicentral distance (86.6°). The long period surface waves are often referred to as "mantle waves." (Courtesy of Shan Dou.)



Surface Waves, Figure 2 Particle motion for Rayleigh and Love waves. Rayleigh waves have retrograde particle motion confined to the vertical plane of propagation. Love waves have purely transverse motion in the horizontal plane. (After Bolt, B. A., 1976. *Nuclear Explosions and Earthquakes*. W. H. Freeman.)

horizontally in a direction perpendicular to the propagation direction (SH energy), suffers from the more complex data processing required, and from higher levels of long period background noise on the horizontal component records, due primarily to the effect of atmospheric pressure variations, inducing ground tilts. Love waves, unlike Rayleigh waves, can exist only in a heterogeneous medium (i.e., a layered medium).

The most striking property of surface waves is their dispersive character, with, in general, waves of longer periods traveling faster, directly reflecting the increase of elastic velocity with depth in the earth's crust and mantle. Indeed, studies of crustal and upper mantle structure progressed rapidly in the 1950s and early 1960s, as the tools developed to measure group and phase velocity dispersion and interpret them in terms of layered mantle and crust models (e.g., Ewing et al., 1957).

In a spherical earth, the theory of generation and propagation of surface waves is intimately linked to that of Earth's free oscillations, leading to the distinction of various modes. The lowest mode is termed fundamental mode, followed by higher modes, or overtones. Fundamental mode Love and Rayleigh waves are generally well separated from other seismic phases on the seismograms, and are well excited by shallow, crustal earthquakes, while higher modes (i.e., overtones) travel at higher group velocities and appear as compact packets of mixed overtones. They are better excited by deeper earthquakes (Figure 3).

Surface waves recorded at teleseismic distances contain information about both the characteristics of the earthquake source and the structure of the earth's crust and mantle along the source-station path. Separating the two effects has been one of the long-standing challenges faced by seismologists.

Group and phase velocity

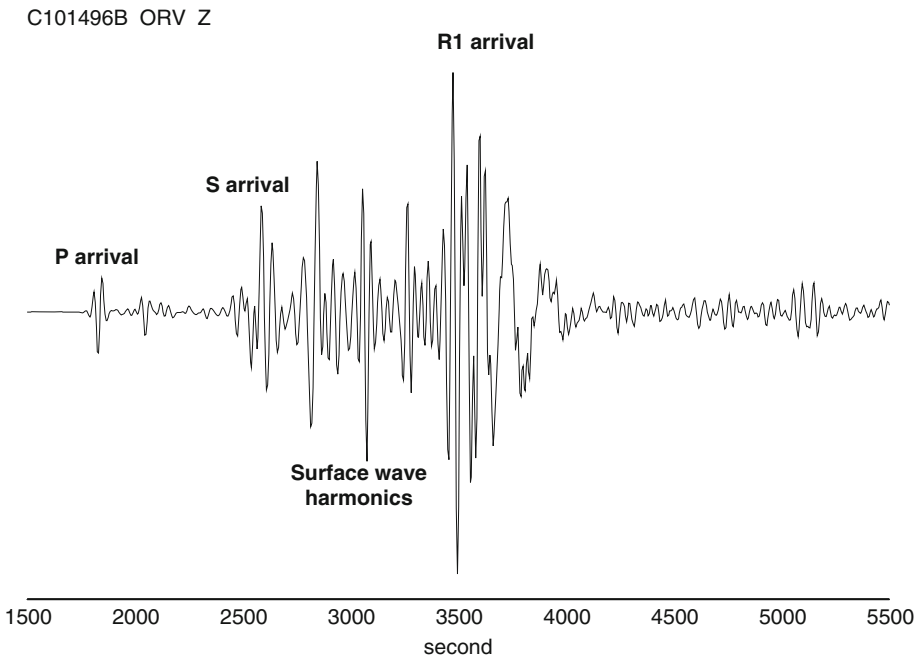
The velocity of propagation of surface waves can be described in terms of either group velocity or phase velocity. Phase velocity is the velocity of a specific frequency component of the wave train. Phase velocity derives from the expression of a plane wave of wavenumber k, and angular frequency ω, propagating in the direction x, for example, in one dimension:

$$f(x,t) \sim \exp i(\omega t - kx) \tag{1}$$

It is defined as the velocity of the phase of the component of frequency ω:

$$C(\omega) = \omega/k \tag{2}$$

where ω is the angular frequency and k is the wavenumber. Frequency is related to period of the wave, T, by $\omega = 2\pi/T$, while $k = 1/\lambda$ where λ is wavelength.



Surface Waves, Figure 3 Example of vertical component record for the October 14, 1996 Solomon Islands earthquake (depth = 19 km, *Mw* = 6.7) at BDSN station ORV showing the arrivals of multiply reflected body wave phases forming a higher-mode Rayleigh wave train in front of the fundamental mode (R1). The Airy phase, corresponding to the group velocity minimum around 230 s, is well visible at the end of the R1 train. The seismogram has been bandpass filtered with cutoff frequencies at 35 and 400 s. Courtesy of Yuancheng Gung.

Group velocity is the velocity with which the energy travels. The group velocity is derived from considering a wave train containing multiple frequencies:

$$u(\omega, t) = \int A(\omega) \exp i(\omega t - kx) d\omega \quad (3)$$

and finding the point (x,t) where the phase is stationary:

$$d(\omega t - kx)/d\omega = 0 \quad \text{hence} \quad t = x dk/d\omega \quad (4)$$

This leads to the definition of group velocity U as:

$$U(\omega) = d\omega/dx \quad (5)$$

In general, at any given period, the group velocity of surface waves in the earth will be less than the phase velocity. The minimum in the group velocity curve is termed Airy phase, a major feature of many surface wave records. Note that waves with periods shorter than the period of the Airy phase will exhibit reverse group velocity dispersion, with the shorter periods traveling faster. Phase velocity, on the other hand, always increases with period.

Surface wave generation and propagation

The theoretical formulation for the excitation of surface waves and normal modes of the earth, an eigenvalue problem, was developed in the 1960s (e.g., Haskell, 1964), much stimulated by the occurrence of the great Chilean earthquake of May 22, 1960 (M 9.5). A computational method, following the original theoretical approach of Saito (1967) based on Runge-Kutta matrix integration method, has long been the main reference for the practical calculation of surface waves and normal modes in laterally homogeneous, elastic, flat or spherical earth models. Today, a widely used approach for spherical geometry, efficient to relatively short periods (10 s), is based on a propagator matrix method in which minors of sets of solutions are used effectively (Woodhouse, 1988). The association of a normal mode formalism to compute dispersion and excitation of surface waves (and complete seismograms), with a moment tensor formalism to describe the earthquake source (e.g., Backus and Mulcahy, 1976) led to the rapid development of source studies based on surface waves in the 1980s.

To obtain the frequency spectrum of a single mode surface wave train from the expression of a seismogram obtained by summation of normal modes in a spherically symmetric earth, one uses Poisson's formula (e.g., Aki and Richards, 1980), which decomposes the modes into infinite trains of propagating surface waves traveling in opposite directions around the earth. In this process, a high frequency approximation is used, in which the phase velocity (see below) of a surface wave is related to the corresponding normal mode frequency by Jeans' formula:

$$C(\omega) = \frac{a\omega_l}{l + 1/2} \quad (6)$$

where l is the angular order of the mode and ω_l its eigenfrequency, and a is the radius of the earth. In this high-frequency approximation, surface waves propagate along the great circle path between the epicenter and the station and are sensitive to structure only along this great circle.

The spectrum of a single mode propagating surface wave at distance Δ, azimuth θ, and angular frequency ω can be expanded as follows:

$$U(\Delta, \theta, \omega) = U_s(\theta, \omega) \mathrm{S}(\Delta) U_p(\Delta, \theta, \omega) \mathrm{F}(\omega, \theta_0) \mathrm{D}(\omega) \mathrm{I}(\omega) \quad (7)$$

where U_s is the source spectrum, U_p contains propagation effects, I is the instrument response, $\mathrm{S}(\Delta)$ is the geometrical spreading term, and F and D describe the source process.

The propagation term U_p can be expressed as (e.g., Romanowicz, 2002):

$$U_p(\Delta, \theta, \omega) = \frac{1}{(\sin \Delta)^{1/2}} \exp(i\pi/4) \exp(im\pi/2) \times \exp[-i\omega\Delta/C(\omega, \theta)] \exp[-\eta(\omega, \theta)\Delta] \quad (8)$$

where m denotes the number of polar passages and $C(\omega,\theta)$, $\eta(\omega,\theta)$ are, respectively, the average phase velocity and attenuation coefficient along the source-station path.

On the other hand, the source term $U_s(\theta,\omega)$ can be expressed as a linear combination of the moment tensor elements M_{ij} of the source.

Structure studies using surface waves

Fundamental mode studies (dispersion)

Fundamental mode surface waves are well suited for the study of crust and upper mantle elastic structure, which can be deduced from their group and/or phase dispersion properties. They allow the sampling of vast areas of the globe that are otherwise devoid of seismic stations, such as the oceans.

Many early studies documented the correlation of seismic velocity variations with surface tectonic features, using regional measurements of phase and group velocities of fundamental mode Love and Rayleigh waves in the period range 20–100 s (e.g., Knopoff, 1972) or at longer periods, reaching deeper into the mantle (e.g., Toksöz and Anderson, 1966). Since the early 1970s, the computation of group velocity has relied on the multiple filtering technique introduced by Dziewonski et al. (1969). This approach, later perfected as the FTAN method (e.g., Lander, 1989), involves two steps. In the first step, an "energy diagram" is formed by windowing the seismogram over group arrival time and filtering over an array of specified center frequencies. The resulting amplitudes are then plotted as a function of time and frequency and their values contoured. The group velocity curve is

obtained by tracing the loci of maximum amplitude as a function of period. In a second step, the time domain seismogram is filtered using multiple filters centered on the group velocity curve (e.g., Figure 4).

Phase velocity is obtained from the phase Φ of the Fourier spectrum of a dispersed wave train that has been corrected for the contribution of the source and the instrument:

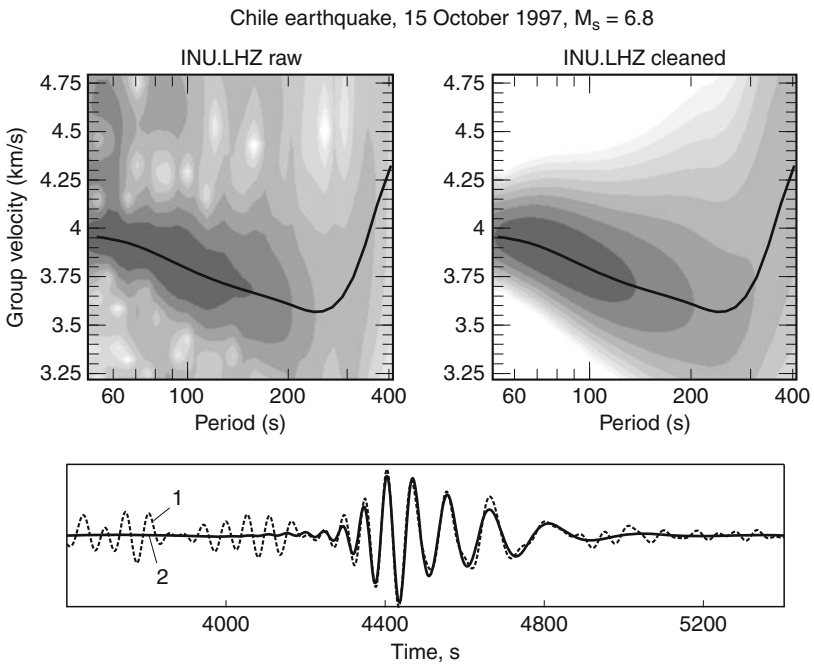
$$C(\omega) = \frac{X}{t_0 - [\Phi - N - (m/4) - 1/8]/\omega} \qquad (9)$$

where t_0 is the start time of the Fourier window with respect to the event's origin time, m is the number of polar passages, and N is an integer arising from the 2π indeterminacy of the phase. This integer is determined first at long periods to obtain reasonable values of phase velocity compatible with well-constrained global earth models. The phase velocity curve is then successively continued to shorter periods at fixed N, so as to obtain a smooth curve. This can become a problem at periods shorter than 30 s, where small variations in phase velocity correspond to rapid cycling of the phase.

The contribution of the earthquake source to the phase needs to be known accurately for the computation of phase velocity, less so for group velocity. This can be circumvented by making measurements using the "two-station method," in which dispersion is measured between two stations approximately aligned with the epicenter on a great circle, thus eliminating the common source phase. If, in addition, the two stations are located within a relatively homogeneous geological province, such measurements are called "pure path" and lead directly to the determination of elastic velocity structure beneath that province. This method has been extended to "many stations" to infer structure beneath an array of stations spanning a geologically homogeneous region.

Noise cross-correlation methodology

Recently, a radically different methodology has been introduced in seismology. The medium's "Green's function" between two seismic stations can be extracted from cross-correlations of records at both stations over long time periods, without the need for using earthquake data (see *Seismic, Ambient Noise Correlation*). These "noise cross-correlations" have been particularly effective for the measurement of the fundamental mode dispersion between two stations in and around the microseismic frequency band (5–30 s), where the background seismic noise is the most energetic. It has been shown that this approach leads to results that are as good as those obtained using standard methods, in regions where seismicity is strong and provides good illumination of the structure using "standard" approaches (e.g., Shapiro et al., 2005).



Surface Waves, Figure 4 Example of group velocity dispersion diagrams obtained for the fundamental Rayleigh wave mode on the vertical record of the Chile Ms 6.8 earthquake of October 15, 1997, at Geoscope station INU, using the FTAN multiple filtering approach. (**a**) Dispersion diagram before filtering; (**b**) same after filtering. The group velocity dispersion curve obtained follows the maxima of energy as delineated by the gray scale contours. (**c**) Corresponding time domain seismogram before (*broken line*) and after (*solid line*) variable filtering. The bandpass is 50–400 s. (Reprinted from Romanowicz 2002.)

This has opened up new horizons for the study of regional crustal structure, especially in seismically quiet regions. There remain some theoretical issues to be resolved, in particular regarding biases introduced by the nonuniform distribution of noise sources around the study region, and the possibility of extracting the overtone part of the Green's function. This is because most of the noise sources that contribute to the reconstruction are located near the earth's surface (see Sect. The earth's background noise spectrum – hum and microseisms) and thus preferentially excite the fundamental mode. Nevertheless, the methodology is developing at a rapid pace.

Global studies of structure using surface waves

With the advent of digital recording in the mid-1970s and the expansion of global digital long-period and later broadband networks (see *Seismological Networks*), the processing of the relatively long time series needed to measure surface waves became much easier and opened the way, in the 1980s to large-scale and global studies of upper mantle structure.

Global studies first proceeded according to a regionalization scheme, in which it was assumed that the depth variation of seismic velocities is the same across each tectonic province. These studies confirmed and extended to longer periods (and hence larger depths) early results on the age dependence of structure in the oceans (e.g., Romanowicz, 2002; see also *Earth's Structure, Upper Mantle*). The constraint of regionalization was soon relaxed, replaced by expansion of lateral heterogeneity into a global basis of spherical harmonics. This revealed, in particular, that the correlation of structure with tectonics disappears below depths of 200–300 km (e.g., Ekström et al., 1997). Dispersion measurements have focused primarily on the fundamental mode, generally in the period range 35–150 s, and thus are generally limited in resolution to depths shallower than ~300–400 km. This approach has led to several generations of global phase velocity maps at discrete frequencies, with increasing spatial resolution (e.g., Figure 5). In these studies, lateral variations of dispersive properties are now well resolved down to wavelengths of ~1,000–2,000 km. At the long period end, the period range is limited by the difficulty of separating consecutive wave trains, and at the short period end, by the increased complexity of surface wave propagation in the strongly heterogeneous crust and uppermost mantle, resulting in lateral refractions and multipathing. The latter are not taken into account in the simple high frequency, great circle propagation assumptions underlying the construction of dispersion maps.

Inversion of surface wave dispersion data

Inversion of surface wave dispersion data generally involves two steps. By combining measurements at a given frequency for many different paths crossing a particular region, or, at longer periods ($T >\sim 40$ s) over the entire globe (e.g., Figure 5), dispersion maps can be obtained by a standard linear inversion procedure, yielding point by point dispersion curves over the region considered.

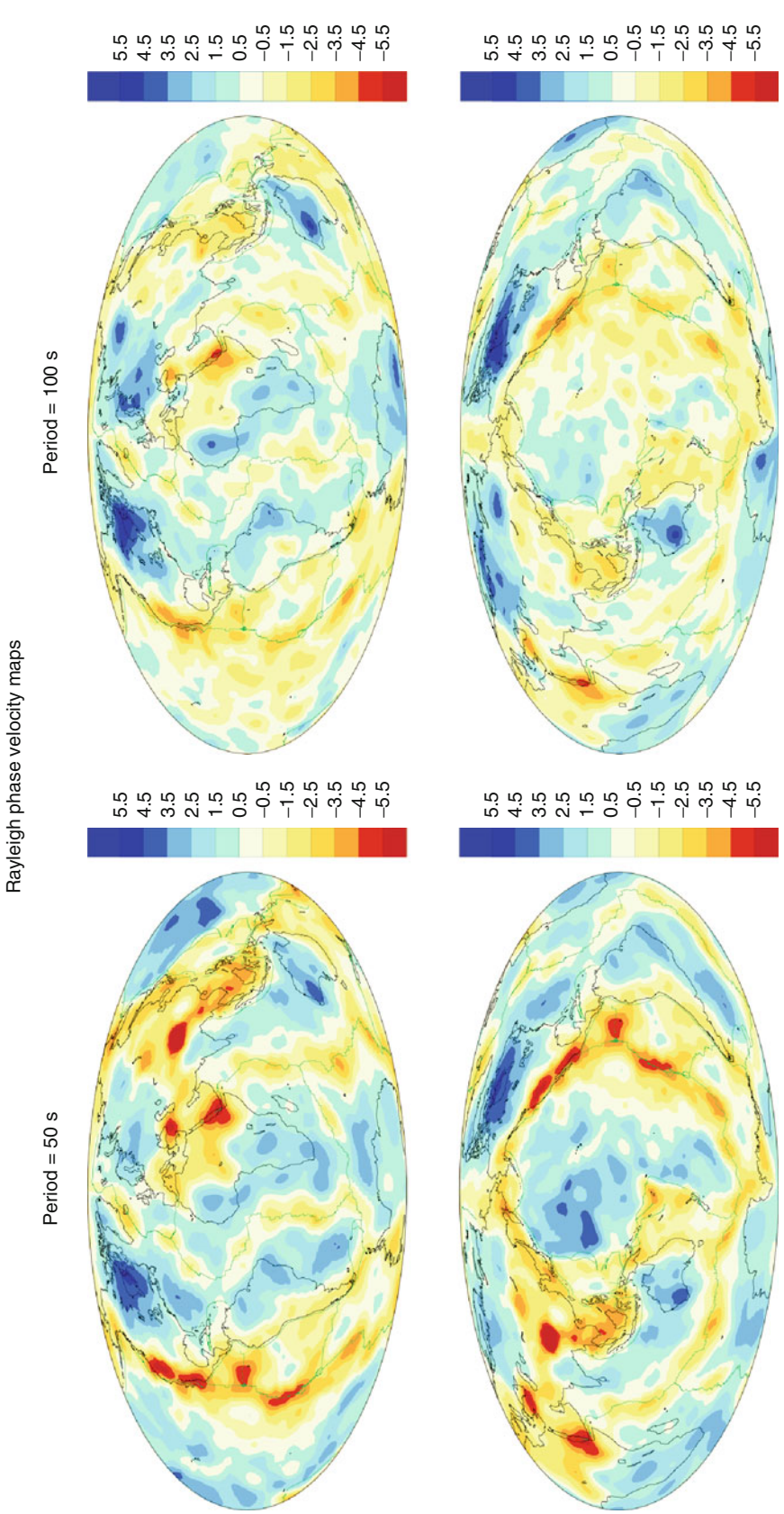
These dispersion curves can then be inverted to obtain the local variation of elastic parameters with depth, by matching observed dispersion curves to theoretical curves calculated for layered earth models, using inverse theory (e.g., Thurber and Ritsema, 2007; see also *Inverse Theory, Linear*). The models thus obtained suffer from some degree of nonuniqueness, as the problem is under-determined. In particular, surface waves cannot by themselves resolve strong gradients of velocity with depth, so that, in general, some a priori constraints are applied, such as fixing the depth of the Moho, or limiting the extent of low velocity layers. A major advance was the theoretical contribution of Backus and Gilbert (1970), which made it possible to calculate the resolving kernels and errors associated with the various least squares solutions of the inverse problem. Surface wave dispersion is most sensitive to shear velocities, so that, in general, a priori conversion factors between shear velocity and compressional velocity, on the one hand, and density, on the other, are assumed.

Crustal corrections

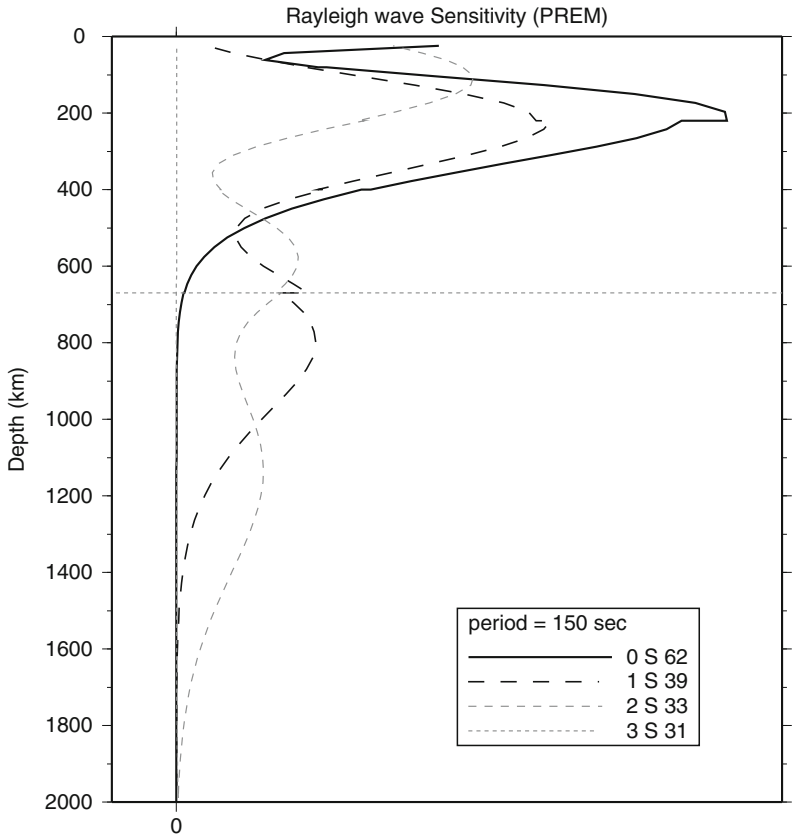
A significant issue regarding the inversion of long period fundamental mode surface waves (40–200 s) for upper mantle structure is that of crustal corrections. Indeed, surface waves are sensitive to shallow structure even at long periods, but not sensitive enough for inversion, unless more difficult to measure short periods (down to ~10 s) are included. Until recently, most studies simply performed crustal corrections in the framework of linear perturbation theory. However, Montagner and Jobert (1988) showed that the effect of the large variations in crustal thickness is nonlinear, and proposed a more accurate correction procedure based on a tectonic regionalization. Developing accurate crustal models worldwide (see *Earth's Structure, Continental Crust*) remains a challenge for large-scale surface wave inversions for structure of the upper mantle, especially in the era of numerical computation of waveforms.

Inversion using overtones

While they are well separated in the time domain from other mode branches, and therefore well suited for frequency domain single mode analysis techniques, fundamental mode surface waves have several shortcomings: at intermediate periods (~20–150 s) their sensitivity to structure below about 200 km is poor, whereas longer period mantle waves, which reach down to the top of the upper mantle transition zone, have poor spatial resolution. In any case, resolving structure in the transition zone, which is also poorly sampled by body waves, requires the inclusion of higher-mode surface waves, whose sensitivity is larger at these depths (e.g., Figure 6). They are also a powerful tool for investigating structures where low velocity zones may be present (e.g., Kovach



Surface Waves, Figure 5 Examples of global Rayleigh wave phase velocity maps at periods of 50 s (*left*) and 100 s (*right*), highlighting the diminishing lateral variations of structure with period (i.e., with depth). These maps were obtained using the dataset of Ekström et al. (1997). (Courtesy of Guy Masters.)



Surface Waves, Figure 6 Sensitivity with depth of Rayleigh wave fundamental mode (*solid line*) and first two overtones (*broken lines*) at a period of 100 s, computed in the PREM model (Dziewonski and Anderson, 1981). The fundamental mode sensitivity peaks around 200 km whereas the overtones have significant sensitivity down to the top of the lower mantle. Courtesy of Yuancheng Gung.

and Anderson, 1964). In general, however, higher-mode surface waves overlap in the time-frequency domain, and single mode dispersion methods therefore cannot be applied.

In the 1970s, array methods were developed independently by Nolet (1975) and Cara (1978) to extract information from overtones, and applied in the period range 20–100 s to paths across Eurasia and the Pacific Ocean, respectively. These methods require a linear regional array of stations approximately aligned with the epicenter (and not in a nodal direction of the source radiation pattern), in order to separate modes in the (ω,k) domain, where k is the wavenumber. They are limited in application to a few regions of the world with relatively dense, linear arrays. A waveform-based method involving the comparison of observed and synthetic seismograms at a single station was proposed by Lerner-Lam and Jordan (1983). In this approach, branch cross-correlation functions (bccf's) are formed between a particular single-mode synthetic and the observed seismograms and used to invert for an average structure along each source-station path. This approach lends itself to the derivation of secondary observables such as group or phase velocity. The main drawback of this methodology is the contamination of the single mode objective function by interference from other mode branches. More recently, van Heijst and Woodhouse (1997) proposed a "mode-branch stripping" method based on a bccf approach combined with a frequency-stepping procedure, yielding single mode dispersion and amplitudes. A global multimode dispersion dataset assembled in this fashion has been used in the construction of several generations of global shear velocity models (e.g., Ritsema et al., 2004). See also *Earth's Structure, Upper Mantle*.

Full waveform inversions

Extracting single mode dispersion information directly from spectra of windowed wave trains is only effective for the fundamental mode, which is well separated on the seismogram, and even in this case, overlapping with overtone wave trains or, at long periods, other orbits of the fundamental mode, can be a problem. A waveform matching approach is therefore desirable, where observed seismograms are compared to synthetics computed for a given reference earth model, and the model is perturbed to improve the fits of the observed and synthetics in the target time window of observation, which can be

optimized. In particular, such an approach is suitable both for fundamental mode and overtone surface waves.

For many years, the computation of long period synthetic seismograms relied on asymptotic approximations to normal mode first order perturbation theory, as developed in the 1970s. The most widely used approximation is the "path average approximation," introduced for waveform modeling by Woodhouse and Dziewonski (1984). The key steps that led to the various formalisms routinely used today are described in Romanowicz et al. (2008). See also *Seismic, Waveform Modeling and Tomography*.

Working under the assumption of PAVA, Nolet (1990) introduced "the partitioned waveform inversion" (PWI), in which inversion for elastic structure proceeds in two steps, more closely related to the two-step inversion using dispersion data. Path integral parameters are defined and retrieved by nonlinear waveform fitting over each path, and in a second step, inverted linearly for elastic structure. The advantage of this approach is that it reduces the number of parameters to be fit in the nonlinear part of the inversion. The PWI also includes progressive filtering of data starting at the longest periods (see *Seismic Tomography*).

Because it includes only coupling of modes along single dispersion branches, the PAVA is rigorously only suitable for single-mode wave trains such as the fundamental mode. The corresponding sensitivity kernels to structure are 1D kernels, that is, they only depend on the laterally averaged structure in the vertical plane between the source and the receiver. When applied to overtones and body waves, coupling across mode branches needs to be included. For a review of these methodologies, see Romanowicz et al. (2008) and Thurber and Ritsema (2007).

Upper mantle anisotropy from surface wave studies

Early studies of surface wave dispersion showed that, in many regions, it is not possible to find a single isotropic model that satisfies both fundamental mode Love and Rayleigh wave dispersion simultaneously. This provided some of the earliest evidence for the presence of anisotropy in the crust and upper mantle (e.g., Anderson, 1961). This discrepancy can be explained by introducing a transversely isotropic medium with a vertical symmetry axis (radial anisotropy), down to at least 200 km depth. It is in this context that the still widely used Preliminary Reference Earth Model (PREM) was constructed (Dziewonski and Anderson, 1981). Gung et al. (2003) showed that radial anisotropy is also present at depths greater than 200 km beneath the lithospheric roots of stable continents and cratons, and explains discrepancies in lithospheric thickness obtained using surface waves of different polarization, thus reconciling lithospheric thickness obtained from seismology with estimates from other geophysical and xenolith data.

Surface wave dispersion also varies with azimuth of propagation, an indication of "azimuthal anisotropy" (e.g., Forsyth, 1975). The global azimuthal variations of Rayleigh and Love wave dispersion at long periods (100–250 s) were first mapped by Tanimoto and Anderson (1985), who showed that the fast direction appears to correlate with flow directions in the mantle. Montagner and Tanimoto (1991) developed the first global model of radial and azimuthal anisotropy in the upper mantle based on fundamental mode surface wave data. While some questions remain about trade-offs between lateral heterogeneity and anisotropy in this type of inversion, Montagner and collaborators have shown that they can explain their datasets with fewer parameters when azimuthal anisotropy is considered than when it is ignored.

To first order, asymptotically, the azimuthal variation of phase velocity (Love or Rayleigh waves) is of the form (Smith and Dahlen, 1973):

$$C(\omega,\theta) = A_0 + A_1\cos(2\theta) + A_2\sin(2\theta) + A_3\cos(4\theta) + A_4\sin(4\theta) \tag{10}$$

where θ is the azimuth of the wavenumber vector defined clockwise from north. The coefficients $A_i(\omega)$ depend linearly on the elastic tensor at every point in the medium. Theoretical expressions and how they can be applied for crust and upper mantle studies are reviewed in Romanowicz (2002) and Montagner (2007). Because Rayleigh waves are sensitive both to shallow crustal and deeper mantle anisotropy, it is important to use a wide frequency range to resolve the depth dependence of anisotropy using surface waves. Montagner and Nataf (1988) introduced "vectorial tomography." By combining radial and azimuthal dispersion terms, under the assumption that the material possesses a symmetry axis (orthotropic medium), they showed that the 3D model can be described using seven elastic parameters (plus density): the five parameters describing transverse isotropy, and two angles describing the orientation in space of the axis of symmetry. Combining long period waveforms and SKS splitting measurements, Yuan and Romanowicz (2010) showed that azimuthal anisotropy provides a powerful tool to detect layering in the continental lithosphere.

Using different seismological tools, including surface wave overtones (e.g., Trampert and van Heijst, 2002), it has also been suggested that azimuthal anisotropy is present in the transition zone. To date, these results are not consistent with each other and still controversial.

Other diagnostic effects of anisotropy in long period surface waves are waveform anomalies caused by Rayleigh-Love coupling, which generates "quasi-Love" waves on vertical components and "quasi-Rayleigh" waves on transverse (e.g., Park and Maupin, 2007). See also *Seismic Anisotropy*.

Effects of scattering and the Born approximation

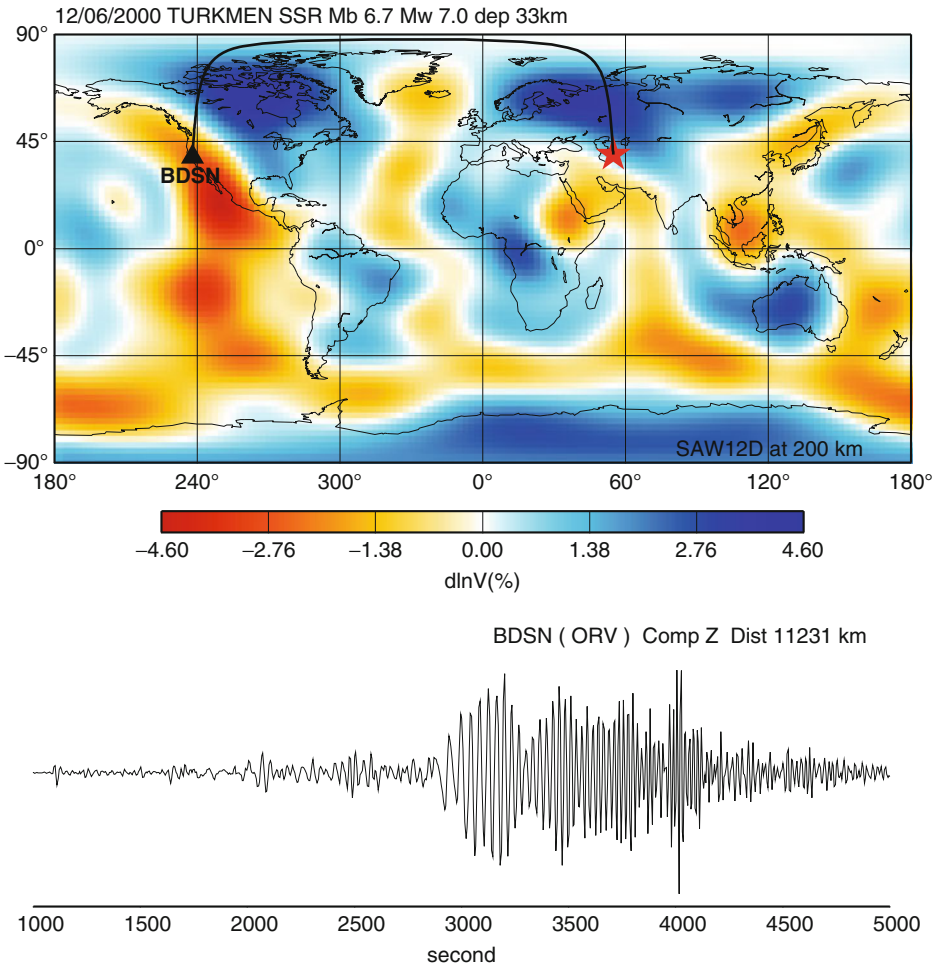
Until recently, most regional and global models of upper mantle structure derived from surface waves were based

on the standard "path average" approximation (PAVA). This is valid only if the wavelength of lateral variations of structure is long with respect to that of the surface waves considered. And it is not strictly appropriate for overtones.

Many observations indicate that lateral heterogeneity is strong enough to cause departures from PAVA. For example, 20 s surface waves sensitive to shallow crustal structure consistently show multipathing (e.g., Capon, 1970 or Figure 7). At longer periods ($T > 100$ s), later arriving trains often show larger amplitudes than the ones preceding them which cannot be explained by lateral variations of attenuation.

Several approaches have been developed to account for these effects and exploit them to obtain better constraints on lateral variations of structure. Woodhouse (1974) introduced the concept of local modes. These are the surface wave modes of a laterally homogeneous model, which locally has the depth distribution of the laterally varying model. If the medium is laterally smooth, the local mode branches propagate as independent wave trains, and their dispersion is modified according to the evolution of the local modes. However, if lateral variations are sharp (e.g., in the presence of a structural discontinuity such as an ocean-continent boundary), the coupling of the local modes cannot be neglected and its strength depends on the width of the structural transition zone (Kennett, 1972). For a recent review see Romanowicz (2002). Kennett (1992) derived a formalism for mode-coupling in 2D slowly laterally varying structures, which was extended to the 3D scattering case, in Cartesian coordinates (e.g., Bostock and Kennett, 1992).

In the framework of scattering theory, the use of tools based on the single-scattering Born approximation has gained popularity in the last 5 years. Born scattering is well suited for inversion since the scattered wavefield



Surface Waves, Figure 7 Example of vertical component record at BDSN station ORV for an *Mw* 7.0 earthquake in Turkmenistan showing multipathing of 20–30 s surface waves. The record has been bandpass filtered between 0.001 and 0.03 Hz. The great circle path is shown on the top plot with a shear wave tomographic model in the background. It arrives to the station from the north, following the coast and along major structural boundaries in the crust that are likely the cause of the multipathing. (Courtesy of Yuancheng Gung.)

depends linearly on structural perturbations (e.g., Snieder, 1988). The use of Born kernels is widely advocated to take into account the effects on dispersion of off-great circle sensitivity of surface waves (e.g., Zhou et al., 2006). However, the relative merits of single-scattering Born sensitivity kernels versus simpler averaging kernels, such as "fat rays" is still debated (e.g., Boschi et al., 2006).

Born is also a poor approximation in the case of strong heterogeneity, or when the region of scattering is large (e.g., Friederich et al., 1993).

Most recently, the introduction to global seismology of the Spectral Element Method (Komatitsch and Vilotte, 1998), which provides an accurate representation of the teleseismic wavefield in arbitrary 3D media up to the second surface wave orbit, has provided a new powerful tool for the study of Earth structure, and is progressively being implemented for waveform modeling of crust and upper mantle structure at the local (e.g., Tape et al., 2010), regional (Fichtner et al., 2010), and global scale (Lekic and Romanowicz, 2010), focusing, so far, on the relatively long period surface waves. These studies show that the use of a more accurate wave propagation theory results in the improved retrieval of amplitudes of lateral variations of shear wave velocity, and in particular, significantly lower values in low velocity zones. For applications at the global scale, this promising new tool presents several challenges: a considerable computational cost, and the necessity of accurate representation of the 3D structure in the crust, a topic that undoubtedly will be addressed in the coming years.

Surface wave attenuation

Surface wave measurements provide the primary constraints on attenuation structure in the crust and uppermost mantle. In the absence of perturbing effects due to scattering and focusing, the amplitude spectrum of a single mode wave train can be written as:

$$A(\omega) = A_0(\omega)\exp(-\eta(\omega)X) \tag{11}$$

where X is the epicentral distance in kilometers and $A_0(\omega)$ represents the amplitude at the source. The attenuation coefficient $\eta(\omega)$ is related to the quality factor Q through (e.g., Aki and Richards, 1980):

$$\eta(\omega) = \frac{\omega}{2C(\omega)Q(\omega)}$$

where C is the phase velocity.

Regional studies of amplitudes of fundamental mode surface waves in the period range 5–100 s and of Lg waves have long established the presence of large variations of Q correlated with tectonic provinces and in particular with the age of the oceans, and with time elapsed since the latest tectonic activity on continents. For recent reviews, see Mitchell (1995) and Romanowicz and Mitchell (2007). Recently, attenuation structure in the vicinity of the East Pacific Rise and in the back-arcs and wedges of subduction zones obtained from a combination of surface wave and body wave data have revealed highly attenuating zones underlying volcanic arcs, down to the slab surface. See also *Seismic, Viscoelastic Attenuation*.

Lateral variations in attenuation at long periods can be an order of magnitude larger than those in elastic velocity. Progress in constraining global 3D anelastic structure of the upper mantle has been slow because of the inherent difficulty of measuring attenuation in the presence of focusing and scattering effects that can be as large as anelastic ones and depend strongly on the short-wavelength details of the elastic structure. The long wavelength 3D attenuation models obtained thus far indicate that lateral variations in attenuation in the first 200–250 km of the mantle are correlated with tectonics. In the transition zone, the long wavelength pattern shifts to one correlated with hotspots and with the structure at the base of the mantle, with high attenuation overlying the two large low shear velocity provinces under Africa and in the Pacific (e.g., Romanowicz and Mitchell, 2007). Progress in resolving finer details of attenuation structure awaits the routine implementation of numerical methods for the computation of long period seismograms to fully account for focusing effects.

Source studies using surface waves

Fundamental mode surface wave spectra contain information about the source moment tensor, source depth (centroid), source process time, and, under favorable circumstances for very large earthquakes, source directivity. Tsai and Aki (1971) first showed that the amplitude spectrum of Rayleigh waves contained the signature of source depth, in the form of a "hole" in the spectrum in the period range 10–100 s, which appears at a period depending on depth, and which also depends on the source mechanism.

In order to correct for propagation, different approaches need to be taken depending on the size of the earthquake and the period range considered. At very long periods ($T > 180$ s) and for large earthquakes ($M > 6.5$), propagation effects can be accounted for approximately using a spherically symmetric reference earth model and ellipticity corrections. At shorter periods, and for smaller earthquakes, corrections on individual source-station paths need to be known much more accurately (for a review, see Romanowicz, 2002). In the last 15 years, the availability of increasingly accurate global 3D tomographic models of the upper mantle has made it possible to extend the period range to shorter periods (down to ~120 s) and to efficiently make use of methodologies, based on time-domain waveform inversion, that are no longer restricted to the fundamental mode, thus providing more accurate estimation of the source depth. Such a waveform approach was first introduced by Dziewonski et al. (1981), who combined waveforms of mantle waves at periods greater than 120 s with overtone waveforms at periods greater than 80 s. This forms the basis of a now routine procedure that serves to construct the widely used Harvard centroid moment tensor (CMT) catalog (Now "global CMT catalog").

The earth's background noise spectrum: hum and microseisms

The presence of background peaks in the frequency range 2–7 mHz, corresponding to the fundamental spheroidal free oscillations of the earth, in the absence of large earthquakes, was discovered recently by Japanese scientists (e.g., Kobayashi and Nishida, 1998). It was later determined that the vertical component long period seismic noise spectrum was made up of Rayleigh waves (Nishida et al., 2002) and that the main source of the "hum" was located in the oceans (Rhie and Romanowicz, 2004), resulting from nonlinear interactions between the atmosphere (oceanic storms), the oceans, and the solid earth. The "hum" has also been detected on the horizontal components, corresponding to Love waves.

The largest background seismic noise is found in the period range 2–20 s and corresponds to the so-called microseismic peak. Microseisms were studied extensively in the 1950s and determined to be surface waves generated in the oceans. There are two types of microseisms: "primary" and "secondary." Secondary microseisms are the strongest. Also called "double-frequency microseisms," they are generated at frequency f by a nonlinear mechanism involving two opposing ocean waves of frequency $f/2$ (Longuet-Higgins, 1950). In the last decade, there has been renewed interest in the study of microseisms, in the context of the use of noise cross-correlations for the study of Earth structure, as seen above. While it has been proposed that the earth's low frequency hum generation mechanism is related to that of microseisms, no common explanation has yet been found for the vertical and horizontal hum generation. See also *Seismic Noise*.

Summary

Surface waves propagate along the surface of the earth with exponentially decreasing energy with depth, with waves of longer period reaching greater depths. They contain most of the elastic energy generated by earthquakes at periods greater than 20 s and are dispersive. There are two types of surface waves, Rayleigh and Love waves, with different polarization properties. The fundamental mode surface waves are well separated from other energy arrivals in the time domain and provide the best constraints to date on continental scale and global scale structure. They play a particularly important role in constraining the upper mantle structure in the ocean basins where few seismic stations have been installed. While the dispersion of fundamental mode surface waves is a classical tool for the investigation of upper mantle structure, higher modes, or overtones provide constraints at transition zone depths (400–700 km) and deeper. They require more sophisticated analysis tools because they cannot be distinguished from each other readily on the seismogram. Waveform inversion approaches are increasingly favored to handle both fundamental mode and overtone interpretation, and hold increasing promise as numerical methods for the computation of the seismic wavefield in arbitrary 3D structures are being implemented, and can account for complex scattering and focusing effects accurately. In addition to isotropic shear velocity, surface waves provide constraints on the distribution of polarization and azimuthal anisotropy in the upper mantle, as well as on anelastic attenuation. When path effects have been corrected for, surface waves provide robust constraints on the source depth and radiation pattern, as expressed by its moment tensor.

The earth's continuous background noise consists primarily of surface waves, at least at periods longer than 5 s. Notable is the microseismic noise peak around 6–7 s. Most of this energy is generated in the oceans by nonlinear interactions involving wind-driven ocean waves, and the seafloor. Recently, a method based on the cross-correlation of noise records has gained popularity and has been applied successfully to resolve crust and uppermost mantle structure. It is particularly useful in seismically quiet regions.

Bibliography

Aki, K., and Richards, P. G., 1980. *Quantitative Seismology, Theory and Methods*. San Francisco: W.H. Freeman.

Anderson, D. L., 1961. Elastic wave propagation in layered anisotropic media. *Journal of Geophysical Research*, **66**, 2953–2963.

Backus, G., and Gilbert, F., 1968. The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, **16**, 169–2050.

Backus, G., and Mulcahy, M., 1976. Moment tensors and other phenomenological descriptions of seismic sources I- continuous displacements. *Geophysical Journal of the Royal Astronomical Society*, **46**, 341–371.

Boschi, L., Becker, T. W., Soldati, G., and Dziewonski, A. M., 2006. On the relevance of Born theory in global seismic tomography. *Geophysical Research Letters*, **33**, L06302, doi:10.1029/2005GL025063.

Capon, J., 1970. Analysis of Rayleigh-wave multipath propagation at LASA. *Bulletin of the Seismological Society of America*, **60**, 1701–1731.

Cara, M., 1978. Regional variations of higher Rayleigh-mode phase velocities: a spatial filtering method. *Geophysical Journal of the Royal Astronomical Society*, **54**, 439–460.

Dziewonski, A. M., and Anderson, D. L., 1981. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, **25**, 297–356.

Dziewonski, A. M., Bloch, S., and Landisman, M., 1969. A new technique for the analysis of transient seismic signals. *Bulletin of the Seismological Society of America*, **59**, 427–444.

Dziewonski, A. M., Chou, A. T., and Woodhouse, J. H., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, **86**, 2825–2852.

Ekström, G., Tromp, J., and Larson, E. W. F., 1997. Measurements and global models of surface wave propagation. *Journal of Geophysical Research*, **102**, 8137–8157.

Ewing, W. M., Jardetsky, W. S., and Press, F., 1957. *Elastic Waves in Layered Media*. New York: McGraw Hill.

Fichtner, A., Kennett, B., Igel, H., and Bunge, H.-P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle. *Earth and Planetary Science Letters*, **290**, 270–280.

Forsyth, D. W., 1975. The early structural evolution and anisotropy of the oceanic upper mantle. *Geophysical Journal of the Royal Astronomical Society*, **43**, 103–162.

Friederich, W., Wielandt, E., and Stange, S., 1993. Multiple forward scattering of surface waves: comparison with an exact solution and Born single-scattering methods. *Geophysical Journal International*, **112**, 264–275.

Gung, Y. C., Panning, M., and Romanowicz, B., 2003. Anisotropy and thickness of the lithosphere. *Nature*, **422**, 707–711.

Haskell, B., 1964. Radiation pattern of surface waves from point sources un a multi-layered medium. *Bulletin of the Seismological Society of America*, **54**, 377.

Kennett, B. L. N., 1972. Seismic waves in laterally heterogeneous media. *Geophysical Journal of the Royal Astronomical Society*, **27**, 301–325.

Knopoff, L., 1972. Observation and inversion of surface-wave dispersion. *Tectonophysics*, **13**, 497–519.

Kobayashi, N., and Nishida, K., 1998. Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature*, **395**, 357–360.

Komatitsch, D., and Vilotte, J.-P., 1998. The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, **88**, 368–392.

Kovach, R. L., and Anderson, D. L., 1964. Higher mode surface waves and their bearing on the structure of the Earth's mantle. *Bulletin of the Seismological Society of America*, **54**, 161–182.

Lander, A. V., 1989. Frequency-time analysis. In Keilis-Borok, V. I. (ed.), *Seismic Surface Waves in a Laterally Inhomogeneous Earth*. Dordrecht: Kluwer Academic, pp. 153–163.

Lekic, V., and Romanowicz, B., 2010. Inferring upper mantle structure by full waveform tomography with the Spectral Element Method. *Geophysical Journal International*, in revision.

Lerner-Lam, A., and Jordan, T. H., 1983. Earth structure from fundamental and higher-mode waveform analysis. *Geophysical Journal of the Royal Astronomical Society*, **75**, 759–797.

Longuet-Higgins, M., 1950. A theory of the origin of microseisms. *Philosophical Transactions of Royal Society London*, **243**, 1–35.

Mitchell, B. J., 1995. Anelastic structure and evolution of the continental crust and upper mantle from seismic surface wave attenuation. *Reviews of Geophysics*, **33**, 441–462.

Montagner, J. P., 2007. Upper mantle structure: global isotropic and anisotropic elastic tomography. In Romanowicz, B., and Dziewonski, A. M. (eds.), *Treatise on Geophysics.* Oxford: Elsevier, Vol. 1, pp. 559–590.

Montagner, J. P., and Jobert, N., 1983. Variation with age of the deep structure of the Pacific Ocean inferred from very long-period Rayleigh wave dispersion. *Geophysical Research Letters*, **10**, 273–276.

Montagner, J. P., and Jobert, N., 1988. Vectorial tomography II. Application to the Indian Ocean. *Geophysical Journal International*, **94**, 309–344.

Montagner, J. P., and Nataf, H. C., 1988. Vectorial tomography – I. Theory. *Geophysical Journal International*, **94**, 295–307.

Montagner, J. P., and Tanimoto, T. 1991. Global upper mantle tomography of seismic velocities and anisotropy. *Journal of Geophysical Research*, **96**(20), 20337–20351.

Nishida, K., Kobayashi, N., and Fukao, Y., 2002. Origin of Earth's ground noise from 2 to 20 mHz. *Geophysical Research Letters*, **29**, 1413, doi:10.1029/2001GL013862.

Nolet, G., 1975. Higher-Rayleigh modes in western Europe. *Geophysical Research Letters*, **2**, 60–62.

Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs. *Journal of Geophysical Research*, **95**, 8499–8512.

Park, J., and Maupin, V., 2007. Theory and observations – wave propagation in anisotropic media. In Romanowicz, B., and Dziewonski, A. M. (eds.), *Treatise on Geophysics*. Oxford: Elsevier, Vol. 1, pp. 323–360.

Rhie, J., and Romanowicz, B., 2004. Excitation of Earth's free oscillations by atmosphere-ocean-seafloor coupling. *Nature*, 431, 552–555.

Ritsema, J., van Heijst, H. J., and Woodhouse, J. H., 2004. Global transition zone tomography. *Journal of Geophysical Research*, **109**, B02302, doi:10.1029/2003JB002610.

Romanowicz, B., 2002. Inversion of surface waves: a review. In Lee, W. H. K. (ed.), *Handbook of Earthquake and Engineering Seismology*, *Part A*. IASPEI, pp. 149–174.

Romanowicz, B., and Mitchell, B. J., 2007. Q in the Earth from crust to core. In Romanowicz, B., and Dziewonski, A. M. (eds.), *Treatise of Geophysics*. Oxford: Elsevier, Vol. 1, pp. 731–774.

Romanowicz, B., Panning, M., Gung, Y., and Capdeville, Y., 2008. On the computation of long period seismograms in a 3D earth using normal mode based approximations. *Geophysical Journal International*, **175**, 520–536.

Saito, M., 1967. Excitation of free oscillations and surface waves by a point source in a vertically heterogeneous Earth. *Journal of Geophysical Research*, **72**, 3689.

Shapiro, N. M., Campillo, M., Stehly, L., Ritzwoller, M. H., 2005. High-resolution surface wave tomography from ambient seismic noise. *Science*, **307**, 1615–1618.

Snieder, R., 1988. Large-scale waveform inversions of surface waves for lateral heterogeneity, 1. Theory and numerical examples. *Journal of Geophysical Research*, **93**, 12055–12065.

Tanimoto, T., and Anderson, D. L., 1985. Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100-250s. *Journal of Geophysical Research*, **90**, 1842–1858.

Tape, C., Liu, J., Maggi, A., and Tromp, J., 2010. Seismic tomography of the Southern California crust based on the spectral-element and adjoint methods. *Geophysical Journal International*, **180**, 433–462.

Thurber, C., and Ritsema, J., 2007. Seismic tomography and inverse methods. In Romanowicz, B., and Dziewonski, A. M. (eds.), *Treatise on Geophysics*. Oxford: Elsevier, Vol. 1, pp. 1323–1360.

Toksöz, M. N., and Anderson, D. L., 1966. Phase velocities of long period surface waves and structure of the upper mantle, 1, Great circle Love and Rayleigh wave data. *Journal of Geophysical Research*, **71**, 1649–1658.

Trampert, J., and van Heijst, H., 2006. Global azimuthal anisotropy in the transition zone. *Science*, **296**, 1297–1299.

Tsai, Y. B., and Aki, K., 1971. Amplitude spectra of surface waves from small earthquakes and underground nuclear explosions. *Journal of Geophysical Research*, **75**, 5729.

van Heist, H., and Woodhouse, J., 1997. Measuring surface-wave overtone phase velocities using a mode-branch stripping technique. *Geophysical Journal International*, **131**, 209–230.

Woodhouse, J. H., 1974. Surface waves in laterally varying structure. *Geophysical Journal of the Royal Astronomical Society*, **37**, 461–490.

Woodhouse, J. H., 1988. The calculation of eigenfrequencies and eigenfunctions of the free oscillations of the earth and the sun. In Doornbos, D. J. (ed.), *Seismological Algorithms*. San Diego: Academic Press, pp. 321–370.

Woodhouse, J. H., and Dziewonski, A. M., 1984. Mapping the upper mantle: three dimensional modelling of the earth structure by inversion of seismic waveforms. *Journal of Geophysical Research*, **89**, 5953–5986.

Yang, Y., and Forsyth, D. W., 2006. Regional tomographic inversion of the amplitude and phase of Rayleigh waves with 2-D sensitivity kernels. *Geophysical Journal International*, **166**, 1148–1160.

Yu, Y., and Park, J., 1994. Hunting for azimuthal anisotropy beneath the Pacific Ocean region. *Journal of Geophysical Research*, **99**, 15399–15421.
Yuan, H., and Romanowicz, B., 2010. Lithospheric layering in the north American craton. *Nature*, **466**, 1063–1069.
Zhou, Y., Nolet, G., Dahlen, F., and Laske, G., 2006. Global upper-mantle structure from finite-frequency surface-wave tomography. *Journal of Geophysical Research*, **111**, B04,304, doi:10.1029/2005JB003677.

Cross-references

Earth's Structure, Continental Crust
Earth's Structure, Upper Mantle
Earthquakes, Source Theory
Free Oscillations of the Earth
Inverse Theory, Linear
Lithosphere, Continental
Propagation of Elastic Waves: Fundamentals
Seismic Anisotropy
Seismic Noise
Seismic Tomography
Seismic, Ambient Noise Correlation
Seismic, Viscoelastic Attenuation
Seismic, Waveform Modeling and Tomography
Seismological Networks