T

T WAVES

Emile A. Okal
Department of Earth & Planetary Sciences, Northwestern University, Evanston, IL, USA

Definition and introduction

T phases are defined as seismic recordings of signals having traveled an extended path as acoustic waves in the water body of the oceans. This is made possible by the "Sound Fixing and Ranging" (SOFAR) channel, a layer of minimum sound velocity acting as a wave guide at average depths of 1,000 m (Okal, 2007). It allows the efficient propagation of extremely small signals over extremely long distances, in practice limited only by the finite size of the ocean basins. The existence of the SOFAR channel results from the dependence of the velocity of sound in water on temperature, pressure and salinity. As a result, the detailed structure of the channel (including the value of the minimum velocity) varies both geographically (mainly with latitude) and seasonally, but at first order, the SOFAR channel can be regarded as a quasi-universal feature of the world's oceans.

At the shoreline of a continent or island, the acoustic wave is converted into a seismic wave which can be recorded by a seismometer or, occasionally, felt by humans. Efficient propagation inside the waveguide requires short wavelengths, in practice frequencies above 3 Hz, and thus T phases generally attenuate fast once converted, although in exceptional geometries they have been detected up to 1,000 km inside stable continental shields. Even though records of T waves were reported as early as 1927, they were not identified as hydroacoustic signals until the late 1940s, when results from antisubmarine warfare during World War II were declassified.

Sources of T waves

T waves can be generated by a diversity of underwater sources, as long as their energy becomes trapped in the SOFAR channel. Most earthquakes in oceanic margins generate strong T phases, since in the presence of a sloping ocean floor, multiple downslope reflections can lead to efficient penetration of the SOFAR (Johnson et al., 1963). However, their generation by earthquakes in abyssal plains deprived of large-scale bathymetry defies the laws of geometrical optics, and remains to this day a major theoretical challenge; it is generally thought that abyssal T phases result from scattering along small random heterogeneities of the ocean floor (Yang and Forsyth, 2003). Also, T waves are occasionally generated by non-shallow earthquakes, even the deepest ones, even though their source is de facto distant from the ocean column. This excitation is made possible by travel through the subducting oceanic slabs which, being cold, propagate high frequencies with limited attenuation and can deliver them to the water interface for conversion to acoustic waves (Okal, 2001a).

Underwater landslides are also T-wave generators. As compared with earthquakes, such sources, which move generally smaller segments of ocean floor over much larger distances, have a considerably longer duration, a lower-frequency spectrum, and thus result in T waves of longer duration but smaller amplitudes (Okal, 2003).

By contrast, explosions at sea, which are intrinsically high-frequency, are particularly efficient sources of T waves; in most instances, they involve a shallow source located inside the SOFAR channel, thus providing optimal feeding of acoustic energy into the channel. In particular, seismic reflection experiments are routinely recorded thousands of kilometers away.

Underwater volcanic eruptions are also regular generators of T waves. Because of the diversity of processes involved during a volcanic sequence, these signals may

Harsh K. Gupta (ed.), *Encyclopedia of Solid Earth Geophysics*, DOI 10.1007/978-90-481-8702-7,
© Springer Science+Business Media B.V. 2011

feature a wide variety of characteristics. The opening of cracks during the formation of magmatic plumbing systems is essentially in the nature of an earthquake, while magmatophreatic explosions at the contact between a lava conduit and the oceanic column are reminiscent of chemical explosions. The steady output of lava into the ocean takes the form of a prolonged noise of low amplitude which may be compared to underwater slumping (Caplan-Auerbach et al., 2001). Finally, the oscillation of magma inside the plumbing system, known to cause volcanic tremor long recorded seismically on subaerial volcanoes, generates harmonic T-wave signals. Supercritical hydrothermal venting during volcanic events has also been documented as the source of monochromatic hydroacoustic signals (Talandier and Okal, 1996).

Among the most remarkable T waves ever reported, signals from the eruption of Macdonald Volcano (29°S; 140°W) on 29 May 1967 led to the eventual discovery of this unchartered underwater seamount, and its interpretation as the active member of the Cook-Austral hotspot chain (Johnson, 1970).

The collision of large icebergs, during which their masses can rub against each other, has also been documented as a source of T waves, which can exhibit a harmonic character during episodes of stick-and-slip friction between tabular icebergs (MacAyeal et al., 2008). Hydroacoustic waves are also generated during the disintegration of large icebergs, and thus could be used to monitor and assess any possible change in climatic conditions (Li and Gavrilov, 2006).

Finally, marine mammals can talk into the SOFAR channel and correspond across hundreds of kilometers, using signals in the range of a few tens of hertz, but specific to each species.

Human perception

T waves of sufficient amplitude can be felt by shoreline populations. For example, the underwater nuclear test WIGWAM on 14 May 1955 off the coast of Southern California was felt in Hawaii and even in Japan, 8,000 km away. Large events such as the Alaskan earthquake of 10 July 1958 and the deep Bolivian shock of 9 June 1994 were felt in Hawaii, and T waves from the 2004 Sumatra earthquake were felt in the Maldives and possibly on Zanzibar; they could have provided a warning of the impending tsunami in these otherwise aseismic areas, but their populations were unprepared.

Use in comprehensive nuclear-test ban treaty monitoring

Because hydroacoustic waves can detect small sources (especially intra-oceanic ones) at great distances, they have been included as one of the four technologies used by the International Monitoring System (IMS) of the Comprehensive Nuclear-Test Ban Treaty. Their spectacular efficiency in propagation allows a full coverage of the all the Earth's oceans using only 11 receiver sites, six of which are hydrophones deployed within the SOFAR channels, and five special seismic stations close to shorelines ("T-phase stations"), instrumented for high-frequency seismic recording in the 50−100 Hz range (Okal, 2001b).

Acoustic thermometry of ocean climate

This project consisted of using acoustic sources in the 57 Hz range fired in the vicinity of Heard Island, South Indian Ocean, a location allowing the illumination of all three major oceans (Munk et al., 1994). Detection at extreme ranges of up to 17,000 km allowed precise measurement of acoustic velocities, and it was envisioned to repeat these experiments over several decades, to monitor any evolution of sound velocities, as a proxy to changes in temperature, possibly in the context of global warming. Unfortunately, the project had to be halted in the late 1990s under pressure from the environmental lobby.

Quantification

Like other seismic phases, quantitative measurements of T waves can provide information on source parameters. However, absolute measurements (especially of amplitudes) at seismic receivers are difficult to interpret due to the complexity and site specificity of the conversion processes. In addition, the high-frequency nature of the phase results in amplitude saturation at relatively low magnitudes. In this context, quantification of T waves has emphasized duration of wavetrains over amplitude, and relative measurements under common or comparable receiver conditions. *Okal et al.* (2003) have introduced the concept of *T*−phase energy flux, *TPEF*, reminiscent of the evaluation of seismic energy from body waves, and which should scale with seismic moment M_0 for regular seismic sources. Any deviation in the ratio $\Gamma = TPEF/M_0$ identifies anomalous source properties; in particular, slow earthquakes exhibit deficient Γ, a property with some potential in tsunami warning.

The comparison of duration and amplitude of T wavetrains can efficiently discriminate between earthquakes and explosions at sea, based on the strong differences in the scaling of their source processes (Talandier and Okal, 2001), although small earthquakes located inside steep volcanic edifices remain a challenge in this respect.

A few geophysical applications

As most of their path is hydroacoustic, T phases propagate at a velocity close to that of sound in water, 1.5 km/s; that gives them late arrivals on seismic records, hence their name "T" for "third". Such slow velocities allow a remarkable resolving power for source locations, after accounting for small geographic and seasonal variations in SOFAR velocities, and for the contribution of the converted [seismic] segments of the phase.

In this context, the superior detection capabilities of T phases have been used to improve considerably our coverage of low-magnitude seismicity in the oceanic

environment, and especially our understanding of small-scale processes at mid-oceanic ridges, notably in conjunction with swarms of volcanic activity (Dziak et al., 1995).

The slowness of T waves also allows to very precisely beam a receiving array (e.g., triangles of hydrophones at sites of the International Monitoring System [IMS]) to back-track the source of T phases, or even of individual fragments of their wavetrains; this procedure was used to obtain a detailed source tomography of the great 2004 Sumatra-Andaman event (e.g., Tolstoy and Bohnenstiehl, 2005).

Conclusion

Because of their high-frequency character, T waves provide a valuable complement to more traditional phases in the field of body-wave seismology. Their incorporation into the IMS allows their more systematic recording and opens up a number of new research opportunities.

Bibliography

Caplan-Auerbach, J., Fox, C. G., and Duennebier, F. K., 2001. Hydroacoustic detection of submarine landslides on Kilauea volcano. *Geophysical Research Letters*, **28**, 1811–1813.

Dziak, R. P., Fox, C. G., and Schreiner, A. E., 1995. The June-July 1993 seismo-acoustic event at CoAxial segment, Juan de Fuca Ridge: Evidence for a lateral dike injection. *Geophysical Research Letters*, **22**, 135–138.

Johnson, R. H., 1970. Active submarine volcanism in the Austral Islands. *Science*, **167**, 977–979.

Johnson, R. H., Northrop, J., and Eppley, R., 1963. Sources of Pacific T phases. *Journal of Geophysical Research*, **68**, 4251–4260.

Li, B., and Gavrilov, A. N., 2006. Hydroacoustic observation of Antarctic ice disintegration events in the Indian Ocean, Proc. First Australasian Acoustical Societies Conference, pp. 479−484, Christchurch.

MacAyeal, D. R., Okal, E. A., Aster, R. C., and Bassis, J. N., 2008. Seismic and hydroacoustic tremor generated by colliding icebergs. *Journal of Geophysical Research*, **113**(F03), 1–10. F03011.

Munk, W. H., Spindel, R. C., Baggeroer, A., and Birdsall, T. G., 1994. The Heard Island feasibility test. *Journal of the Acoustical Society of America*, **96**, 2330–2342.

Okal, E. A., 2001a. "Detached" deep earthquakes: Are they really? *Physics of the Earth and Planetary Interiors*, **127**, 109–143.

Okal, E. A., 2001b. T − phase stations for the international monitoring system of the comprehensive nuclear-test ban treaty: A global perspective. *Seismological Research Letters*, **72**, 186–196.

Okal, E. A., 2003. T waves from the 1998 Papua New Guinea earthquake and its aftershocks: Timing the tsunamigenic slump. *Pure and Applied Geophysics*, **160**, 1843–1863.

Okal, E. A., 2007. The generation of T waves by earthquakes. *Advances in Geophysics*, **49**, 1–65.

Okal, E. A., Alasset, P.-J., Hyvernaud, O., and Schindelé, F., 2003. The deficient T waves of tsunami earthquakes. *Geophysical Journal International*, **152**, 416–432.

Talandier, J., and Okal, E. A., 1996. Monochromatic T waves from underwater volcanoes in the Pacific Ocean: ringing witnesses to geyser processes? *Bulletin of the Seismological Society of America*, **86**, 1529–1544.

Talandier, J., and Okal, E. A., 2001. Identification criteria for sources of T waves recorded in French Polynesia. *Pure and Applied Geophysics*, **158**, 567–603.

Tolstoy, M., and Bohnenstiehl, D. R., 2005. Hydroacoustic constraints on the rupture duration, length, and speed of the great Sumatra-Andaman earthquake. *Seismological Research Letters*, **76**, 419–425.

Yang, Y., and Forsyth, D. W., 2003. Improving epicentral and magnitude estimation of earthquakes from T phases by considering the excitation function. *Bulletin Seismological Society of America*, **93**, 2106–2122.

Cross-references

Seismic Monitoring of Nuclear Explosions
Seismology, Monitoring of CTBT

THERMAL STORAGE AND TRANSPORT PROPERTIES OF ROCKS, I: HEAT CAPACITY AND LATENT HEAT

Christoph Clauser
Institute for Applied Geophysics and Geothermal Energy, E.ON Energy Research Center, RWTH Aachen University, Aachen, Germany

Synonyms

Specific heat capacity; Thermal capacity; Volumetric heat capacity

Definition

Specific heat capacity c. Physical property defining the amount of sensible heat that can be stored in or extracted from a unit mass of rock per unit temperature increase or decrease, respectively. Isobaric and isochoric specific heat capacities are defined at constant pressure and volume, respectively; dimension: J kg^{-1} K^{-1}.
Thermal capacity (*also: volumetric heat capacity*) ρ c. The product of isobaric specific heat capacity and density. Physical property defining the amount of sensible heat that can be stored in or extracted from a unit volume of rock per unit temperature increase or decrease, respectively; dimension: J m^{-3} K^{-1}.

Thermal storage properties

The thermal regime of the Earth is defined by its heat sources and sinks, the heat storage and transport processes, and their corresponding physical properties. The storage properties are discussed below. The transport properties, thermal conductivity and thermal diffusivity, are dealt with in this volume in the companion chapter "*Thermal Storage and Transport Properties of Rocks, II: Thermal conductivity and diffusivity*."

Heat can be stored as sensible heat and enthalpy of transformation. In the Earth, sensible heat is defined by the heat capacity of rocks, and the enthalpy of transformation by their latent heat of fusion.

Heat is transmitted within the Earth mainly by diffusion (in the steady state: conduction), advection, and radiation. Generally, heat diffusion dominates heat radiation within the lithosphere of the Earth, at temperatures below

about 1 000 K. For sufficiently large flow rates, convection-driven heat advection, however, can be equally or even more efficient than diffusion. The heat diffusion–advection equation for a porous medium is:

$$\frac{\partial(\rho c_P T)}{\partial t} = \nabla \cdot \left(\underbrace{\lambda \nabla T}_{q_{dif}} - \underbrace{(\rho c_P)_f T \mathbf{v}}_{q_{adv}} \right) + \underbrace{A}_{\text{generation}}, \quad (1)$$

where T is temperature in K, t time in s, ρ density in kg m^{-3}, c isobaric specific heat capacity in J kg^{-1} K^{-1}, λ thermal conductivity in W m^{-1} K^{-1}, **v** specific fluid discharge (volume flow rate per unit cross section) in m s^{-1}, A (radiogenic) heat generation rate in W m^{-3}; subscript "f" refers to fluid properties. Diffusive bulk storage and transport properties for heat in rocks governing this equation are thermal capacity ρc_P, also referred to as volumetric heat capacity, and thermal conductivity λ. Advection is governed by the thermal capacity and specific discharge of the flowing fluid, $(\rho c_P)_f$ and **v**, respectively.

Heat advection does not require very large flows for becoming as or more efficient than heat diffusion. This is often the case in sedimentary basins (see, e.g., Clauser et al. (2002) for a literature review). But fluid-driven heat advection may be also important in crystalline rocks and on a crustal scale (e.g., Clauser, 1992). The non-dimensional Péclet and Nusselt numbers, Pe and Nu, are ratios of specific heat flows q_{adv} and q_{dif} (Equation 1) indicating the efficiency of (fluid flow driven) advective heat transport versus heat conduction for given flow geometries. For instance, assuming flow of magnitude v over a distance L and a temperature difference $T_1 - T_0$ one obtains:

$$\begin{aligned} Pe &= \frac{q_{adv}}{q_{dif}} = \frac{(\rho c)_f v (T_1 - T_0)}{\lambda (T_1 - T_0)/L} = \frac{(\rho c)_f v L}{\lambda}; \\ Nu &= \frac{q_{adv} + q_{dif}}{q_{dif}} = \frac{q_{adv}}{q_{dif}} + 1 = Pe + 1. \end{aligned} \quad (2)$$

Thus advection or diffusion (in the steady state: conduction) dominate for $Pe > 1$ or $Pe < 1$, respectively (in terms of the Nusselt number for $Nu > 2$ or $Nu < 2$, respectively). At temperatures above 1 000 K, heat is propagated increasingly as radiated electromagnetic waves, and heat radiation begins to dominate diffusion (see *Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity*, this volume).

Heat capacity

Heat can be stored and delivered as sensible heat or as latent heat required or liberated by phase changes. This and the next two paragraphs are concerned with sensible heat. Following this, latent heat will be discussed.

Sensible heat capacity C is defined as the ratio of heat ΔQ required to raise the temperature of a mass M of rock by ΔT. For each molecule of mass m, this temperature increase requires an energy of (f/2) k ΔT, where f is the number of degrees of freedom of the molecule and k = 1.380 650 4(24) × 10^{-23} J K^{-1} is Boltzmann's constant. For a body of mass M, a temperature increase of ΔT requires an energy of ΔQ = (M/m) (f/2) k ΔT. Thus the heat capacity of the body at constant volume is:

$$C_V = \frac{\Delta Q}{\Delta T} = \frac{f}{2} k \frac{M}{m} = \frac{f}{2} k N_A = \frac{f}{2} R \left(J K^{-1} \right), \quad (3)$$

where Avogadro's number N_A = 6.022 141 79(30) × 10^{23} mol^{-1} equals the number of molecules or atoms in an amount of substance of 1 mol and R = k N_A = 8 314 472(15) J mol^{-1} K^{-1} is the molar Gas constant (numerical values for N_A, R, k, and all other physical constants used in this chapter are from CODATA, 2006). For solids, f = 6, corresponding to the three degrees of freedom of potential and kinetic lattice vibration energy in each space direction. Accordingly, the heat capacity of one mole of substance, the molar heat capacity at constant volume is constant:

$$C_{V,mol} = 3 k N_A = 3R = 24.94 \ (J \ mol^{-1} K^{-1}) \quad (4)$$

where N_A = 6.022 141 79(30) × 10^{23} mol^{-1} is Avogadro's number, the number of molecules in one mole of substance, and R = k N_A the molar gas constant. Isobaric heat capacity $C_{P,mol}$ is larger than isochoric heat capacity $C_{V,mol}$ because additional work is required for volume expansion. Both are related by:

$$C_{P,mol} = C_{V,mol} + R = \frac{f+2}{2} R \ (J \ mol^{-1} K^{-1}). \quad (5)$$

With $C_{V,mol}$ from Equation 4 and assuming, as above, f = 6 this yields:

$$\begin{aligned} C_{P,mol} &= 3R + R = 4R \\ &= 33.26 \ (J \ mol^{-1} K^{-1}), \text{ or:} \\ C_{P,mol} &- C_{V,mol} = R. \end{aligned} \quad (6)$$

Equation 4, the *Dulong–Petit law*, is satisfied well for heavy elements. In contrast, molar heat capacities of lighter elements remain below this limiting value, the lower the temperature the smaller $C_{V,mol}$.

Below the Debye temperature, Θ_D, heat capacity varies with temperature. Θ_D tends to zero as T^3 as absolute temperature approaches zero. Θ_D falls in the range 85 K and 450 K for most substances and 200 K and 1 000 K for most minerals (Stacey and Davis, 2008; Table 1). Therefore, heat capacity in the Earth can be well explained by classical Debye theory, in particular in the mantle and except for a thin crustal layer near the Earth's surface. There are, however, exceptions such as beryllium (Θ_D = 1 440 K) and diamond ($\Theta_D \approx$ 1 800 K). These are caused by the so-called freezing of vibrational or rotational degrees of freedom, which cannot absorb heat any more at low temperature. Therefore heat capacity tends to zero close to absolute zero.

Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat, Table 1 Debye temperature Θ_D and mass number A of selected elements (Kittel, 2004)

Element	Mass number A	Θ_D (K)
Carbon	6	2 230
Aluminum	13	428
Silica	14	645
Titanium	22	420
Chromium	24	630
Manganese	25	410
Iron	26	470
Nickel	28	450
Copper	29	343
Zinc	30	327
Silver	47	215
Cadmium	48	209
Tin	50	200
Tantalum	73	240
Tungsten	74	400
Platinum	78	240
Gold	79	165
Lead	82	105

Isobaric and isochoric specific heat capacity

Specific heat capacity c of a substance is defined as heat capacity C related to unit mass:

$$c = \frac{\Delta Q}{M\,\Delta T} = \frac{f}{2}\frac{k}{m} = \frac{f}{2}\frac{k}{A_r\,m_u}\left(J\,kg^{-1}K^{-1}\right), \tag{7}$$

where $m_u = 1.660\,538\,782(83) \times 10^{-27}$ kg is the atomic mass constant, defined as $^1/_{12}$ of the atomic mass of the carbon isotope ^{12}C, and A_r the atomic mass of a substance relative to m_u. Isobaric specific heat capacity c_P is larger than isochoric specific heat capacity c_V because additional work is required for volume expansion. Their ratio, the adiabatic exponent, is:

$$c_P/c_V = (f+2)/f. \tag{8}$$

Alternatively, isobaric specific heat capacity c_P can be expressed by enthalpy H(T,P) = E + P V, a state function of temperature and pressure, where E, P, and V are internal energy, pressure, and volume, respectively. In a closed system, the change in internal energy dE is the sum of the change in heat dQ and the work dW delivered: dE = dQ + dW. If we only consider volume expansion work, dW = −P dV, the change in enthalpy dH becomes:

$$\begin{aligned} dH(T,P) &= dE + pdV + VdP = dQ + VdP \\ &= \left(\frac{\partial H}{T}\right)_P dT + \left(\frac{\partial H}{P}\right)_T dP. \end{aligned} \tag{9}$$

Comparing coefficients, we obtain:

$$\frac{dQ}{dT} = \left(\frac{\partial H}{\partial T}\right)_P \doteq c_P. \tag{10}$$

Thus, Equation 10 defines isobaric specific heat capacity c_P as the first derivative of enthalpy with respect to temperature. Comparison of Equations 7 and 10 shows that both expressions are equivalent for dQ = ΔQ/M, and the isobaric enthalpy change ΔH is equal to the specific heat content ΔQ/M.

Isobaric and isochoric specific heat capacity are related to compressibility β = ΔV/(VΔP) and its inverse, incompressibility or bulk modulus K = VΔP/ΔV, by $c_P/c_V = \beta_T/\beta_S = K_S/K_T$ (e.g., Stacey and Davis, 2008), where subscripts T and S refer to isothermal and adiabatic conditions, respectively. Inserting the thermodynamic relation $\beta_T = \beta_S + \alpha^2\,T/(\rho\,c_P)$ (e.g., Birch, 1966) between isothermal and adiabatic compressibility yields the relative difference between isobaric and isochoric specific heat capacity:

$$c_P/c_v = 1 + \alpha\gamma T, \tag{11}$$

where α = ΔV/(VΔT) is the volume expansion coefficient,

$$\gamma = \frac{\alpha K_S}{\rho c_P} = \frac{\alpha K_T}{\rho c_V}, \tag{12}$$

the dimensionless Grüneisen parameter, and ρ density. Inserting the expressions for α and K into Equation 12 yields:

$$\gamma = \frac{1}{\rho c_P}\frac{\Delta V}{V\,\Delta T}\frac{V\Delta P}{\Delta V} = \frac{\Delta P}{\rho c_P \Delta T}. \tag{13}$$

Thus the Grüneisen parameter γ is the relative pressure change in a material heated at constant volume.

For solids, i.e., f = 6, the absolute difference between isobaric and isochoric specific heat capacity follows from Equations 11 and 8:

$$c_P - c_V = \frac{K_T\alpha^2 T}{\rho} = \frac{3K_S\alpha^2 T}{4\rho}. \tag{14}$$

For crustal rocks (γ = 0.5; α = 20 μK^{-1}; T < 10^3 K; ρ = 2 600 kg m^{-3}; K_S < 75 GPa (Dziewonski and Anderson, 1981; Stacey and Davis, 2008)), the difference between isobaric and isochoric specific heat capacity is less than 9 J kg^{-1} K^{-1} or 1% according to Equations 14 and 11, respectively. Thus, the distinction between isobaric and isochoric specific heat capacity is negligible for crustal rocks at temperatures below 1 000 K. However, it need be made for mantle rocks. From here on, "specific heat capacity" will always refer to isobaric specific heat capacity, denoted simply by the letter c without the subscript "P."

This classical treatment of heat capacity is sufficient for temperatures above the Debye temperature. In the Earth, temperature exceeds the Debye temperature everywhere except in the crust (Stacey and Davis, 2008). Therefore, in experiments at room temperature and atmospheric pressure, we observe deviations from the values predicted by Equations 3–14, which are based on the classical Dulong–Petit theory. The lower the temperature, lighter the element, and stronger the lattice bonding become, the larger are these deviations. Clearly, interpretation of heat

capacity below the Debye temperature is beyond classical mechanics and requires quantum mechanical treatment. This is, however, beyond the scope of this text and interested readers are referred to standard physics textbooks (e.g., Tipler and Mosca, 2007). Therefore heat capacity at crustal temperatures should not be calculated from Equations 3 and 14 but rather be measured or calculated from appropriate, quantum mechanical equations.

Čermák and Rybach (1982) compiled data on isobaric specific heat capacity for different rock-forming minerals and different igneous, metamorphic, volcanic, and sedimentary rocks as well as the corresponding variations with temperature.

Measuring techniques

Specific heat capacity c can be measured directly or derived as the isobaric derivative of enthalpy H with respect to temperature. Specific heat capacity of rocks varies with temperature, pressure, porosity, and saturants. Accordingly, in situ values may deviate from laboratory data according to temperature, pressure, and type and content of pore fluid.

Numerous steady-state and transient calorimetric methods are available for measuring specific heat capacity. The most popular are mixing or drop calorimeters and heat flux differential scanning (DSC) calorimeters. The first method yields an absolute value; the second one is a comparative method. All of these methods and their details are discussed in the literature (e.g., Hemminger and Cammenga, 1989; Brown, 2001; Haines, 2002) to which interested readers are referred. The isobaric enthalpy change (or specific heat content) ΔH of solids may be determined by the method of mixtures using a Bunsen-type calorimeter, in which the unknown isobaric enthalpy change of a sample relative to a base temperature, e.g., 25 °C, is compared to the corresponding known isobaric enthalpy change of platinum (Kelley, 1960; Somerton, 1992).

Calculated heat capacity

When no direct measurements can be performed, the isobaric enthalpy change and specific heat capacity of rocks can be calculated according to Kopp's law, Equation 15, as the arithmetic mean of the individual mineralogical and fluid contributions weighted by the volume fractions n_i of the N individual phases relative to total rock volume:

$$\Delta H = \sum_{i=1}^{N} n_i \Delta H_i; \quad c = \sum_{i=1}^{N} n_i c_i; \quad 1 = \sum_{i=1}^{N} n_i. \qquad (15)$$

Based on data for various minerals (e.g., Kelley, 1960; Berman and Brown, 1985; Somerton, 1992; Waples and Waples, 2004), the isobaric enthalpy change ΔH or specific heat capacity c can be computed from Equation 15 for any rock consisting of an arbitrary number of minerals with given volume fractions.

Temperature dependence

Derived from measured variation of isobaric enthalpy change ΔH with temperature of various oxides, Kelley (1960) suggested a second-order polynomial for fitting ΔH from which $c_P = (\partial H/\partial T)_P$ can be easily calculated. Somerton (1992) and Clauser (2006) report ΔH and c values of various rock-forming oxides and pore fluids. An alternative approach is fitting heat capacity measured at different temperatures directly to polynomials of various degrees (e.g., Maier and Kelley, 1932; Berman and Brown, 1985; Fei and Saxena, 1987; Holland and Powell, 1996; Robertson and Hemingway, 1995). Waples and Waples (2004) provide a discussion of the various approaches. The polynomial proposed by Berman and Brown (1985),

$$C_{P,mol} = k_0 + k_1 T^{-0.5} + k_2 T^{-2} + k_3 T^{-3} (T \text{ in K}), \qquad (16)$$

works over a large temperature range and yields no values incompatible with the Dulong–Petit law for high temperatures. Table 2 lists values for the coefficients k_0–k_3 in Equation 16 determined from fits of heat capacity of selected minerals measured at different temperatures.

As an alternative, Waples and Waples (2004) propose a statistical approach describing the general temperature dependence of all rocks and minerals, which can be rescaled easily for individual rocks and minerals. To this end, measured specific heat capacity c_P was normalized by the corresponding normalizing value $c_{P,n}$ at 200 °C (473.15 K), a temperature at or near which data was available. The resulting polynomial regressions yielded much better coefficients of determination R^2 for data measured on nonporous rock ($R^2 = 0.93$) than for those measured on minerals ($R^2 = 0.62$) while the trends were similar. The regression on the combined data for minerals and nonporous rocks yields an expression for the normalized specific heat capacity of a mineral or nonporous rock at arbitrary temperature T with a coefficient of determination $R^2 = 0.65$:

$$c_{P,n}(T) = 0.716 + 1.72 \times 10^{-3} T - 2.13 \times 10^{-6} T^2 + 8.95 \times 10^{-10} T^3, (T \text{ in } ^\circ\text{C}). \qquad (17)$$

Equation 17 can be rescaled for any mineral or nonporous rock at any temperature T_2 provided a value $C_P(T_1)$ measured at temperature T_1 is available, for instance from any of the compilations of Berman and Brown (1985), Fei and Saxena (1987), Berman (1988), Holland and Powell (1996), or Robertson and Hemingway (1995):

$$c_P(T_2) = c_P(T_1) \frac{c_{P,n}(T_2)}{c_{P,n}(T_1)}. \qquad (18)$$

Additionally, Waples and Waples (2004) consider the variation of specific heat capacity with lithology, where interested readers find a specific discussion regarding coals of different carbon content or maturity.

Mottaghy et al. (2005) used a second-order polynomial in temperature to fit the variation of isobaric specific heat

Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat, Table 2 Coefficients for calculating isobaric molar heat capacity $C_{P,mol}$ (J mol^{-1} K^{-1}) (From Equation 16, Berman, 1988)

Mineral	Chemical composition	k_0 (J mol^{-1} K^{-1})	$k_1 \times 10^{-2}$ (J mol^{-1} K$^{-½}$)	$k_2 \times 10^{-5}$ (J mol^{-1} K^{1})	$k_3 \times 10^{-7}$ (J mol^{-1} K^{2})	T (K)
Albite	$NaAlSi_3O_8$	393.64	−24.155	−78.928	107.064	250–1 373
Almandine	$Fe_3Al_2Si_3O_{12}$	573.96	−14.831	−292.920	502.208	420–997
Andalusite	Al_2SiO_3	236.48	−11.029	−75.268	93.644	254–1 601
Anorthite	$CaAl_2Si_2O_8$	439.37	−37.341	0.0	31.702	292–1 373
Calcite	$CaCO_3$	178.19	−16.577	−4.827	16.660	257–1 200
Cordierite	$Mg_2Al_4Si_3O_{18}$	954.39	−79.623	−21.173	−37.021	256–1 652
Corundum	Al_2O_3	155.02	−8.284	−38.614	40.908	250–2 300
Dolomite	$CaMg(CO_3)_2$	328.48	−25.544	−46.885	79.038	250–650
Enstatite (ortho)	$MgSiO_3$	166.58	−12.066	−22.706	27.915	254–1 273
Fayalite	Fe_2SiO_4	248.93	−19.239	0.0	−13.910	255–1 370
Forsterite	Mg_2SiO_4	238.64	−20.013	0.0	−11.624	253–1 807
Hematite	Fe_2O_3	146.86	0.0	−55.768	52.563	258–1 757
Kaolinite	$Al_2Si_2O_5(OH)_4$	523.23	−44.267	−22.443	9.231	256–560
Kyanite	Al_2SiO_3	262.68	−20.014	−19.997	−6.318	252–1 503
Lime	CaO	58.79	−1.339	−11.471	10.298	250–1 176
Magnesite	$MgCO_3$	162.30	−11.093	−48.826	87.466	288–750
Magnetite	Fe_3O_4	207.93	0.0	−72.433	66.436	257–1 825
Muscovite	$KAl_3Si_3O_{10}(OH)_2$	651.49	−38.732	−185.232	274.247	257–967
Periclase	MgO	61.11	−2.962	−6.212	0.584	250–1 798
Potassium feldspar	$KAlSi_3O_8$	381.37	−19.411	−120.373	183.643	250–997
Quartz	SiO_2	80.01	−2.403	−35.467	49.157	250–1 676
Sillimanite	Al_2SiO_5	256.73	−18.827	29.774	25.096	253–1 496
Sphene	$CaTiSiO_5$	234.62	−10.403	−51.183	59.146	255–1 495
Spinel	$MgAl_2O_4$	235.90	−17.666	−17.104	4.062	256–1 805
Talc	$Mg_3Si_4O_{10}(OH)_2$	664.11	−51.872	−21.472	−32.737	250–639
Wollastonite	$CaSiO_3$	149.07	−6.903	−36.593	48.435	251–1 433

capacity with temperature measured on a suite of metasedimentary, volcanic, magmatic, and metamorphic rocks:

$$c_P(T) = A_0 + A_1 T^1 + A_2 T^2 (c_p \text{ in J kg}^{-1}\text{K}^{-1}, 1\,^\circ\text{C} \le T \le 100\,^\circ\text{C}). \quad (19)$$

The average values for the coefficients A_0–A_2 determined from a regression of $c_P(T)$ measured over a temperature range of 1 °C–100 °C on 26 samples from seven boreholes are: $\bar{A}_0 = 0.074717725$ J kg^{-1} K^{-1}; $\bar{A}_1 = 1.862585346$ J kg^{-1} K^{-2}; $\bar{A}_2 = -2510.632231$ J kg^{-1} K^{-3}.

Based on a composition of 30 % quartz, 60 % feldspar (albite), and 10 % phyllosilicates (5% phlogopite, and 5 % annite), Whittington et al. (2009) suggest average "bulk crustal" molar specific heat capacity equations based on end-member mineral data for two temperature ranges, separated by the transition at 846 K (~573 °C) between α- and β-quartz:

$$C_{P,mol}(T) = \begin{cases} 199.50 + 0.0857T - 5.0 \times 10^6 T^2; & T \le 846\text{K} \\ 229.32 + 0.0323T - 47.9 \times 10^{-6} T^2; & T > 846\text{K} \end{cases}; \quad (C_{P,mol} \text{ in J mol}^{-1}\text{K}^{-1}). \quad (20)$$

Assuming an average molar mass of 0.22178 kg mol^{-1}, this yields the variation of isobaric specific heat capacity c_P with temperature shown in Figure 1.

Volumetric heat capacity: thermal capacity

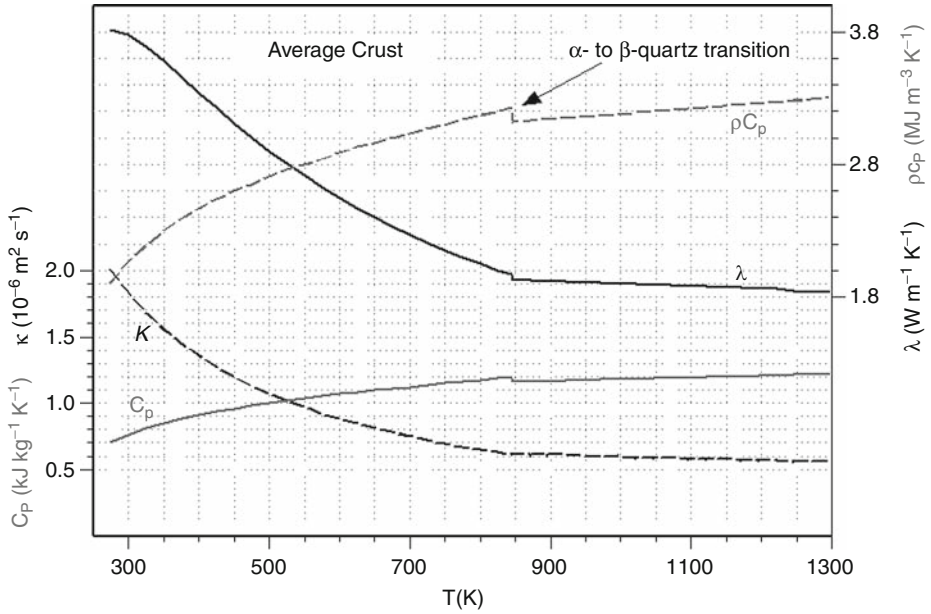
When heat capacity is related to unit volume rather than to unit mass or unit amount of substance, it is referred to as volumetric heat capacity or thermal capacity. It can be calculated as the product of specific heat capacity c and density ρ or as the ratio of thermal conductivity λ and thermal diffusivity κ by

$$\rho c = \lambda/\kappa. \quad (21)$$

Again, Kopp's law yields the rock's bulk thermal capacity $(\rho\, c)_b$ as:

$$(\rho\, c)_b = (1 - \phi)\,(\rho\, c)_s + \phi \sum_{i=1}^{N} S_i\,(\rho c)_i, \quad (22)$$

where ϕ is porosity, $(\rho\, c)_s$ thermal capacity of the rock skeleton, S_i fractional saturation, and $(\rho\, c)_i$ thermal capacity of the i^{th} fluid phase in the pore space. The skeleton thermal capacity itself may be calculated again from Kopp's law for a given mineral assemblage and the corresponding volume fractions of the solid phase from Equation 15. Because of the low density of air and gas – about three orders of magnitude lower than that of water



Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat, Figure 1 Variation of specific heat capacity c_P, the phonon components of thermal diffusivity and thermal conductivity κ_p and λ_p, respectively, and thermal capacity $\rho c_P = \lambda_p/\kappa_p$ with temperature in an average crust according to Equation 20.

and rock – the contribution of the gas phase to thermal capacity can often be ignored. In this case, N = 2 for the fluid phases water and oil or N = 1 for water only. Expressions for the density of various fluids are reported in Clauser (2006). Data on density of various minerals and rocks are listed, e.g., in Wohlenberg (1982a,b) or Olhoeft and Johnson (1989). Using Equations 17 and 18, Waples and Waples (2004) analyzed a substantial collection of density and specific heat capacity data from various authors and transformed specific heat capacity and thermal capacity to a uniform reference temperature of 20 °C (Table 3).

The mean thermal capacity of "impervious" rocks was found at 2.300(46) MJ m^{-3} K^{-1} by Roy et al. (1981). This is acceptably close to the mean of 2.460(65) MJ m^{-3} K^{-1} found by Waples and Waples (2004) for inorganic minerals.

Based on density, specific heat capacity, and thermal conductivity measured at room temperature, Mottaghy et al. (2005) determined thermal capacity as inverse slope of a regression of diffusivity on thermal conductivity according to Equation 21. All values fell well within ±20 % of the average of 2.3 MJ m^{-3} K^{-1} recommended by Beck (1988).

Mottaghy et al. (2008) determined average values for thermal capacity according to Equation 21 for metamorphic and magmatic crystalline rocks as the inverse slope of a linear regression of values of thermal diffusivity versus thermal conductivity measured at temperatures in the range 20 °C–300 °C:

$$\kappa(T) = \frac{\lambda(T)}{m + nT}, (T \text{ in } ^\circ C), \quad (23)$$

Regression of data measured on seven samples collected along a profile crossing the Eastern Alps from north to south and on nine samples from the northern rim of the Fennoscandian Shield near the Kola ultra-deep borehole SG-3 yielded m = 20 66(70) kJ m^{-3} K^{-1}, n = 2.2 (4) kJ m^{-3} K^{-2}, R^2 = 0.97 and m = 2404(91) kJ m^{-3} K^{-1}, n = 3.6(5) kJ m^{-3} K^{-2}, R^2 = 0.92, respectively. This yields a range of thermal capacity for the Alpine and Fennoscandian data of about 2.1 MJ m^{-1} K^{-1}–2.7 MJ m^{-1} K^{-1} and 2.4 MJ m^{-1} K^{-1}–3.5 MJ m^{-1} K^{-1}, respectively, in the temperature range 20 °C–300 °C.

The product of an average density of 2 700 kg m^{-3} for the crust and isobaric specific heat capacity calculated for an average molar mass of 0.22178 kg mol^{-1} according to Equation 20 (Whittington et al., 2009) yields the variation of thermal capacity ρc_P (Equation 21) with temperature shown in Figure 1. Its increase with temperature by about a factor of 1.7 in a temperature interval of 1 000 K demonstrates that the effect of temperature is stronger for the phonon component of thermal diffusivity κ than for phonon thermal conductivity λ_p due to the increase of specific heat capacity. Assuming a constant density throughout the crust implies that the increase and decrease in density due to the increase in pressure and temperature, respectively, partly cancel each other and that these changes are small compared to those of specific heat capacity and thermal diffusivity.

Latent heat

Solidification of magma and melting of rocks as well as freezing and thawing of water in soils or rocks liberates or consumes heat, respectively. The like applies to mineral

phase changes such as those associated with the seismic discontinuities at 410 km, 520 km, and 660 km in the transition zone from 400 km–600 km between the upper and lower mantle. These mineral phases are chemically identical but differ with respect to crystal structure and therefore elastic properties. This is why this transition is seen in the seismic wave field. Phase transitions require a certain pressure P and temperature T, but also a specific relation between these two state variables expressed by the so-called Clapeyron slope $dP/dT = \Delta S(P,T)/\Delta V(P,T)$ (the inverse of the Clausius–Clapeyron equation), where S and V are entropy and volume, respectively. This means that the depth where a certain phase transition occurs varies with the ambient temperature in the crust. Positive and negative values for the Clapeyron slope are associated with exothermic and endothermic reactions, respectively.

Phase changes generally consume or deliver much more latent heat than can be stored or delivered as sensible heat: It requires a temperature increase of more than 500 K to equal by sensible heat the amount of latent heat required to melt 1 kg of granite, and still an increase of more than 80 K to equal by sensible heat the amount of latent heat required to melt 1 kg of sea ice (Table 4).

The *discontinuity at 410 km* is generally associated with the transition in olivine from α-olivine to β-spinel, also named wadsleyite (e.g., Stacey and Davis, 2008). At expected pressure and temperature of 14 GPa and 1,600 K, respectively, corresponding values for the Clapeyron slope vary between 2.9 MPa K^{-1} and 4.8 MPa K^{-1} (Bina and Helffrich, 1994; Katsura et al., 2004; Stacey and Davis, 2008). The *discontinuity at 520 km* is associated with the transition from β-spinel (wadsleyite) into γ-spinel, also named ringwoodite (e.g., Stacey and Davis, 2008). At an expected pressure of 18 GPa, a temperature increase, such as by an ascending plume, would require a higher pressure for this transition according to inferred Clapeyron slopes of 4.0 MPa K^{-1}–5.3 MPa K^{-1} (Helffrich, 2000; Deuss and Woodhouse, 2001). A second transition occurs between garnet and calcium-perovskite ($CaSiO_3$), where the iron in garnet goes into Ilmenite, and its $CaSiO_3$-component into calcium-perovskite. This reaction has a negative Clapeyron slope. The two slopes of different sign may shift the depth for two transitions into opposite directions, which is observed as a splitting of the 520 km discontinuity. The *discontinuity at 660 km* defines the transition into the lower mantle. It is caused by the transition of γ-spinel (ringwoodite) into magnesium-perovskite ($MgSiO_3$) and ferrous periclase (magnesiowüstite, (Fe,Mg)O). At an expected pressure of 23.5 MPa, this endothermic transition is associated with a Clapeyron slope of − 2.8 MPa K^{-1} (Stacey and Davis, 2008).

The latent heat L that corresponds to these additional heat sources and sinks can be elegantly combined with the specific sensible heat capacities of the liquid and solid rock, c_l and c_s, respectively, into an effective bulk specific heat capacity c_{eff}. This effective specific heat capacity then accounts for the entire enthalpy change, including latent heat. In this approach, the latent heat effects are assumed to occur between the solidus and liquidus temperatures T_1 and T_2, respectively. The heat liberated by a solidifying

Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat, Table 3 Typical values or ranges for density ρ, isobaric specific heat capacity c_P, and thermal capacity ρ c_P of selected rocks at 20°C (Waples and Waples, 2004; Petrunin et al., 2004)

Rock	ρ (kg m^{-3})	c_P (J kg^{-1} K^{-1})	ρ c_P (kJ m^{-3} K^{-1})
Albite	2 540–2 560	755–780	1 922–1 991
Amphibole	3 010	700–1 134	2 110–3 410
Anhydrite	2 950–2 960	590–940	1 740–2 780
Anorthite	2 740	800	2 202
Basalt	2 870	880–900	2 526–2 583
Clay	2 680	860	2 300
Coal	1 350	1 300	1 760
Diabase	2 790	731–860	2 040–2 400
Dolomite	2 800	900	2 520
Gabbro	2 970–3 000	650–1 000	1 950–2 970
Gneiss	2 700	770–979	2 080–2 640
Granite	2 620–2 650	600–1 172	1 590–3 070
Gypsum	2 370	1 010	2 390
Limestone	2 760–2 770	680–880	1 880–2 430
Peridotite	2 740–3 190	705–1 005	1 930–3 210
Pyroxenite	3 190–3 240	660–1 000	2 140–3 190
Quartzite	2 640	731–1 013	1 930–2 670
Rock salt	2 160	880	1 900
Sandstone	2 640	775	2 050
Schist	2 770–2 900	790–1 096	2 190–3 180
Serpentinite	2 270–2 540	730–1 005	1 660–2 550
Siltstone	2 680	910	2 449
Slate	2 770–2 780	740–1 113	2 060–3 080
Syenite	2 820	460	1 300
Talc	2 780	1 000	2 780
Tuff	2 750	1 090	3 000

Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat, Table 4 Comparison of isobaric specific heat capacity c_P and latent heat of melting L of granite, basalt, seawater, and freshwater (Stacey and Davis, 2008, supplemented)

	Granite	Basalt	Seawater	Freshwater	Dry air (15°C, 1 atm)
c_P (kJ kg^{-1} K^{-1})	0.83	0.88	3.99	4.18	1.006
L (kJ kg^{-1})	420	420	335	333.55	196

("freezing") liquid phase is obtained by weighting by the volume fractions of liquid and solid phases, ϕ_l and ϕ_s, respectively. The enthalpy change of the rock volume then becomes $dH_{freezing} = (\phi_l\, c_l + \phi_s\, c_s)\, dT + L\, d\phi_l$, and the effective heat capacity c_{eff} is:

$$c_{eff}^{freezing} = \frac{dH}{dT} = \phi_l\, c_l + \phi_s\, c_s + L\, \frac{d\phi_l}{dT}. \quad (24)$$

Conversely, when considering melting the solid phase, the enthalpy change of the rock volume is $dH_{melting} = (\phi_l\, c_l + \phi_s\, c_s)\, dT + L\, d\phi_s$, and the effective heat capacity in this case c_{eff} is:

$$c_{eff}^{melting} = \frac{dH}{dT} = \phi_l\, c_l + \phi_s\, c_s + L\, \frac{d\phi_s}{dT}. \quad (25)$$

Acknowledgments

This contribution benefitted from insightful comments by an anonymous reviewer and editorial help by Dr. Sukanta Roy, Hyderabad.

Bibliography

Beck, A. E., 1988. Methods for determining thermal conductivity and thermal diffusivity. In Hänel, R., Rybach, L., and Stegena, L. (eds.), *Handbook of Terrestrial Heat Flow Density Determination*. Dordrecht: Kluwer, pp. 87–124.

Berman, R. G., 1988. Internally-consistent thermodynamic data for minerals in the system Na_2O-K_2O-CaO-MgO-FeO-Fe_2O_3-Al_2O_3-SiO_2-TiO_2-H_2O-CO_2. *Journal of Petrology*, **29**(2), 445–522.

Berman, R. G., and Brown, T. H., 1985. Heat capacity of minerals in the system Na_2O-K_2O-CaO-MgO-FeO-Fe_2O_3-Al_2O_3-SiO_2-TiO_2-H_2O-CO_2. *Contributions to Mineralogy and Petrology*, **89**, 168–183.

Bina, C. R., and Helffrich, G., 1994. Phase transition Clapeyron slopes and transition zone seismic discontinuity topography. *Journal of Geophysical Research*, **99**(B8), 15853–15860, doi:10.1029/94JB00462.

Birch, F., 1966. Section 7: compressibility; elastic constants. In Clark S. P., Jr. (ed.), *Handbook of Physical Constants.* Rev. ed., Memoir 97, Washington: Geologoical Society of America, pp. 97–173.

Brown, M. E., 2001. *Introduction to Thermal Analysis: Techniques and Applications (Hot Topics in Thermal Analysis and Calorimetry)*. Dordrecht: Kluwer.

Čermák, V., and Rybach, L., 1982. Thermal conductivity and specific heat of minerals and rocks. In Angenheister, G. (ed.), *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology*. Berlin: Springer. New Series, Vol. V(1a), pp. 305–343.

Clauser, C., 1992. Permeability of crystalline rocks. *EOS Transactions of the American Geophysical Union*, **73**(21), 233, 237. (for an update see also: Clauser, C., 2001. *Update of the permeability of crystalline rocks*, Report, Institute for Applied Geophysics and Geothermal Energy, E.ON Energy Research Center, RWTH Aachen University, http://www.eonerc.rwth-aachen.de/go/id/tsm/; retrieved 16 Oct 2010.)

Clauser, C., 2006. Geothermal Energy. In Heinloth, K. (ed.), *Landolt-Börnstein, Group VIII: Advanced Materials and Technologies*. Heidelberg: Springer. Energy Technologies, Subvol. C: Renewable Energies, Vol. 3, pp. 480–595.

Clauser, C., Griesshaber, E., and Neugebauer, H. J., 2002. Decoupled thermal and mantle helium anomalies – Implications for the transport regime in continental rift zones. *Journal of Geophysical Research*, **107**(B 11), 2269, doi:10.1029/2001JB000675.

CODATA (Committee on Data for Science and Technology), 2006. *CODATA Internationally Recommended Values of the Fundamental Physical Constants*, Paris: International Council of Scientific Unions (ICSU), http://physics.nist.gov/cuu/Constants/index.html, retrieved 9 Sept 2010.

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

Dziewonski, A. M., and Anderson, D. L., 1981. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, **25**, 297–356.

Fei, Y., and Saxena, S. K., 1987. An equation for the heat capacity of solids. *Geochimica et Cosmochimica Acta*, **52**(2), 251–254.

Haines, P. (ed.), 2002. *Principles of Thermal Analysis and Calorimetry*, Cambridge, UK: The Royal Society of Chemistry.

Helffrich, G., 2000. Topography of the transition zone seismic discontinuities. *Reviews of Geophysics*, **38**, 141–158.

Hemminger, W. F., and Cammenga, H. K., 1989. *Methoden der Thermischen Analyse*. Berlin: Springer.

Holland, T. J. B., and Powell, R., 1996. An internally consistent thermodynamic data set for phases of petrological interest. *Journal of Metamorphic Geology*, **16**, 309–343.

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-i, 2004. Olivine-wadsleyite transition in the system $(Mg, Fe)_2SiO_4$. *Journal of Geophysical Research*, **109**, B02209, doi:10.1029/2003JB002438.

Kelley, K., 1960. *Contributions to the Data on Theoretical Metallurgy: XIII High-Temperature Heat-Content, Heat-Capacity, and Entropy Data for the Elements and Inorganic Compounds*, U. S. Bureau of Mines Bulletin 584, Washington DC: U. S. Government Printing Office.

Kittel, C., 2004. *Introduction to Solid State Physics*, 8th edn. Hoboken: Wiley.

Maier, C. G., and Kelley, K. K., 1932. An equation for the representation of high temperature heat content data. *Journal of American Chemical Society*, **54**(8), 3243–3246.

Mottaghy, D. C., Schellschmidt, R., Popov, Y. A., Clauser, C., Kukkonen, I. T., Nover, G., Milanovsky, S., and Romushkevich, R. A., 2005. New heat flow data from the immediate vicinity of the Kola super-deep borehole: vertical variation in heat flow confirmed and attributed to advection. *Tectonophysics*, **401**(1–2), 119–142, doi:10.1016/j.tecto.2005.03.005.

Mottaghy, D., Vosteen, H.-D., and Schellschmidt, R., 2008. Temperature dependence of the relationship of thermal diffusivity versus thermal conductivity for crystalline rocks. *International Journal of Earth Sciences*, **97**, 435–442.

Olhoeft, G. R., and Johnson, G. R., 1989. Densities of rocks and minerals. In Carmichael, R. S. (ed.), *Practical Handbook of Physical Properties of Rocks and Minerals*. Boca Raton: CRC Press, pp. 139–176.

Petrunin, G. I., Popov, V. G., and Il'in, I. A., 2004. Conductive heat transfer in plagioclases. *Izvestiya, Physics of the Solid Earth (English Translation)*, **40**(9), 752–759.

Robertson, E. C., and Hemingway, B. S., 1995. *Heat Capacity and Heat Content of Rocks*, Open-file report 95-622. Reston, VA: U. S. Geological Survey.

Roy, R. F., Beck, A. E., and Touloukian, Y. S., 1981. *Thermophysical Properties of Rocks*, In Touloukian, Y. S., Judd, W. R., Roy, R. F., (eds.), *Physical Properties of Rocks and Minerals,* McGraw-Hill/CINDAS Data Series on Material Properties, New York: McGraw-Hill, Vol. II-2, pp. 409–502.

Somerton, W. H., 1992. *Thermal Properties and Temperature Related Behavior of Rock/Fluid Systems*. Amsterdam: Elsevier.
Stacey, F. D., and Davis, P. M., 2008. *Physics of the Earth*, 4th edn. Cambridge: University Press.
Tipler, P. A., and Mosca, G. P., 2007. *Physics for Scientists and Engineers: Extended Version*. Basingstoke: Palgrave Macmillan.
Waples, D. W., and Waples, J. S., 2004. A review and evaluation of specific heat capacities of rocks, minerals, and subsurface fluids. Part 1: Minerals and Nonporous rocks. *Natural Resources Research*, **13**(2), 97–122.
Whittington, A. G., Hofmeister, A. M., and Nabelek, P. I., 2009. Temperature-dependent thermal diffusivity of the Earth's crust and implications for magmatism. *Nature*, **458**, 319–321, doi:10.1038/nature07818.
Wohlenberg, J., 1982a. Density of minerals. In Hellwege, K.-H. (ed.), *Group V: Geophysics Vol 1: Physical Properties of Rocks, Subvol. A*. Berlin: Springer. Landolt-Börnstein, Vol. 1, pp. 66–113.
Wohlenberg, J., 1982b. Density of rocks. In Hellwege, K.-H. (ed.), *Group V: Geophysics Vol 1: Physical Properties of Rocks, Subvol. A*. Berlin: Springer. Landolt-Börnstein, Vol. 1, pp. 113–120.

Cross-references

Geothermal Record of Climate Change
Heat Flow, Continental
Heat Flow, Seafloor: Methods and Observations
Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity

THERMAL STORAGE AND TRANSPORT PROPERTIES OF ROCKS, II: THERMAL CONDUCTIVITY AND DIFFUSIVITY

Christoph Clauser
Institute for Applied Geophysics and Geothermal Energy, E.ON Energy Research Center, RWTH Aachen University, Aachen, Germany

Synonyms

Heat conductivity; Thermal conductivity; Thermal diffusivity

Definition

Thermal conductivity (also: heat conductivity) λ: Physical property governing heat diffusion in the steady state. It defines how much heat flows across a unit cross section of rock along a unit distance per unit temperature decrease per unit time; dimension: $W\,m^{-1}\,K^{-1}$.
Thermal diffusivity κ: Physical property governing transient heat diffusion. It is defined by the ratio of thermal conductivity and thermal capacity, i.e., by the ratio of heat flowing across the face of a unit volume and the heat stored in the unit volume per unit time; dimension: $m^2\,s^{-1}$.

Thermal conductivity

Fourier's law of heat conduction defines the vector of specific heat flow q_i, i.e., heat flow normalized by area, as the product of the thermal conductivity tensor λ_{ij} and the temperature gradient vector $\partial T/\partial x_j$:

$$q_i = -\lambda_{ij}\,\frac{\partial T}{\partial x_j}. \quad (1)$$

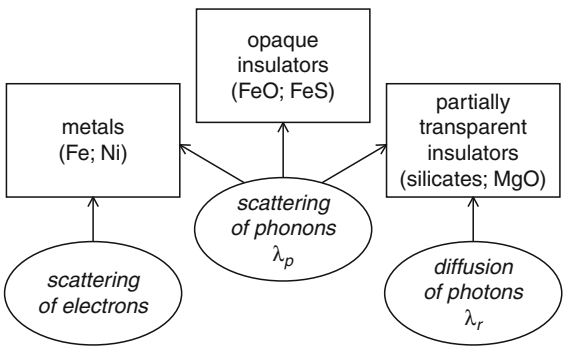
Temperature measurements are usually performed along vertical profiles in boreholes, yielding only the vertical component of the temperature gradient. Thermal conductivity in some rocks is, to a good approximation, isotropic, particularly in volcanic and plutonic rocks. Then heat will flow predominantly vertically, and it is sufficient to consider only the vertical component of Equation 1. In contrast, thermal conductivity of many sedimentary and metamorphic rocks is strongly anisotropic, and lateral heat flow may be significant. Hence, information on anisotropy is often required, demanding laboratory measurements in different directions. Anisotropy occurs over a wide range of scales, from microscopic, over laboratory and macroscopic, to tectonic.

Heat transport in most of the Earth's crust and mantle is diffusive and caused by (1) scattering of quantized lattice vibrations, the phonons, and (2) by diffusive (as opposed to ballistic) radiation of photons. These two processes are described by phonon thermal conductivity λ_p and radiative thermal conductivity, λ_r, respectively, the sum of which is often termed effective thermal conductivity, λ_{eff}. In most polycrystalline material heat radiation starts to dominate at temperatures above 2 500 K. In single crystals and glasses with little or no scattering (e.g., obsidian), however, radiation may become important at much lower temperatures of 500 k–1 000 K. In the metallic core, scattering of electrons provides a third heat transfer mechanism (Figure 1).

The following text deals mostly with phonon thermal conductivity. Radiative heat transfer is discussed in the paragraph on radiative thermal conductivity and in the chapter on thermal diffusivity.

Measuring techniques

Thermal conductivity can be measured in the laboratory on rock samples, i.e., cores or cuttings or in situ either in boreholes or with shallow penetration (3 m–20 m) marine



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 1 Different types of material in the Earth and associated mechanisms of heat transport (Redrawn after Hofmeister et al., 2009).

heat flow probes. There are numerous steady-state and transient techniques available for measuring thermal conductivity, the most prominent being the "divided bar," "needle probe," "optical scanning," and "(laser) flash," all of which are suitable to determine also the anisotropy of thermal conductivity. These methods are described in, for instance, Parker et al. (1961), Kappelmeyer and Haenel (1974), Beck (1988), Davis (1988), Somerton (1992), Schilling (1999), Popov et al. (1999b), Beardsmore and Cull (2001), and Blumm and Lemarchand (2002). Among these techniques, the transient ones are also suitable for determining thermal diffusivity.

As with most other petrophysical properties, in situ thermal conductivity may deviate significantly from laboratory values, even if the effects of temperature, pressure, and pore fluid are accounted for. This scale dependence involves different aspects: In situ measurements represent an average over a much larger rock volume than laboratory measurements performed on small samples, and cannot resolve small-scale variations. A subsequent upscaling may be necessary to identify the appropriate representative elementary volume (REV) for which reasonable transport parameter averages can be defined.

Indirect methods

When no data are available or no direct measurements can be performed, thermal conductivity can be inferred indirectly, either from mineralogical composition and saturating fluids or from correlations with other physical properties. While some of these methods are based on well-defined physical models, others are purely empirical.

Estimation from mineralogical composition and saturating fluids: Thermal conductivity of rocks may be estimated from their mineral and fluid content. Thermal conductivity of minerals varies much less than in rocks, due to their well-defined composition. As the bulk thermal conductivity of porous rocks varies with different saturants, it may be also of interest to know rock thermal conductivity for other saturants than those used in the laboratory measurement. Numerous models based on volume fractions of the individual mineral and fluid phases have been proposed, all with specific advantages and disadvantages: Some overestimate while others underestimate systematically the true bulk thermal conductivity. Most of them are valid only within a specific range of volume fractions (or porosities) and yield unreasonable results outside. To overcome this problem, additional parameters may be introduced in order to incorporate rock structure into a mixing law (see below).

The parallel and series models for thermal resistance of layered media are easy to understand, but have the disadvantage of being rather special cases, applicable mostly to bedded sediments. They correspond to the well-known weighted arithmetic and harmonic means, λ_{ari} and λ_{har}, Equations 2a and b, and are related to heat flowing parallel or perpendicular to bedding, respectively. They define upper and lower limits for all other models, sometimes also referred to as the Voigt upper bound and Reuss lower bound, respectively. Thus they constrain the maximum variance of possible predictions. The arithmetic mean between the two is known as the Voigt-Reuss-Hill average, λ_{VRH} (Equation 2c). It is useful for estimating an effective thermal conductivity, while the arithmetic and harmonic means are used to define a range for possible values.

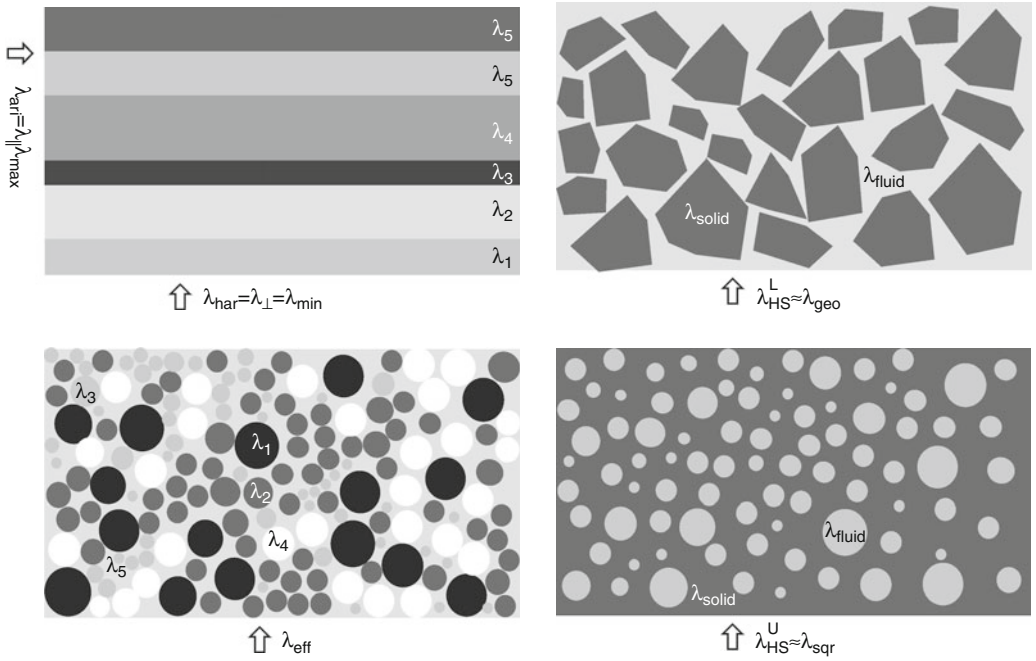
The weighted geometric and square root means λ_{geo} and λ_{sqr} (Equations 2d and e; e.g., Beardsmore and Cull, 2001), respectively, are associated with a mixture of different fluid and mineral phases of unspecified geometry. In many cases, both means successfully explain the data and are therefore widely used for lack of more specific information on the geometric arrangement of the individual volume fractions. In the special case where $N = 2$ and $n_1 = n_2 = 0.5$, the arithmetic, harmonic, and geometric means are related by $\lambda_{geo} = \sqrt{\lambda_{ari}\,\lambda_{har}}$.

Effective medium theory (Bruggeman, 1935) provides the effective medium mean λ_{eff} (Equation 2f) which is useful for macroscopically homogeneous and isotropic rocks consisting of randomly distributed grains and pores.

The upper and lower Hashin-Shtrikman bounds λ^U_{HS} and λ^L_{HS} (Equation 3, Hashin and Shtrikman, 1962), respectively, provide tighter constraints for the predictions of different models other than the arithmetic and harmonic means. The arithmetic average of both defines the Hashin-Shtrikman mean (Equation 2g). Geometrically, the lower Hashin-Shtrikman bound λ^L_{HS} corresponds to a rock model consisting of grains suspended in a fluid and is closely followed by the geometric mixing law. In contrast, the square root law is very close to the upper Hashin-Shtrikman bound λ^U_{HS} and could be related to a well-lithified rock with spherical, fluid-filled pores. Figure 2 illustrates the geometries corresponding to all of these models.

If λ_i denotes the thermal conductivity and n_i the volume fraction of the ith phase relative to the total volume ($1 = \sum n_i$), these seven weighted means are defined by:

$$
\begin{aligned}
&(a)\ \lambda_{\max} = \lambda_{ari} = \lambda_{||} = \sum_{i=1}^{N} n_i\ \lambda_i;\\
&(b)\ \lambda_{\min} = \lambda_{har} = \lambda_{\perp} = \left(\sum_{i=1}^{N} \frac{n_i}{\lambda_i}\right)^{-1};\\
&(c)\ \lambda_{VRH} = \frac{1}{2}\left(\lambda_{||} + \lambda_{\perp}\right);\ (d)\ \lambda_{geo} = \prod_{i=1}^{N} \lambda_i^{n_i};\\
&(e)\ \sqrt{\lambda_{sqr}} = \sum_{i=1}^{N} n_i\ \sqrt{\lambda_i};\ (f)\ \lambda_{eff}^{-1} = \sum_{i=1}^{N} \frac{3\ n_i}{2\ \lambda + \lambda_i};\\
&(g)\ \lambda_{HS} = \frac{1}{2}\left(\lambda^U_{HS} + \lambda^L_{HS}\right);
\end{aligned}
\tag{2}
$$



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 2 Geometrical arrangement of layers, mineral grains, and pores assumed in different models for calculating mean bulk thermal conductivity of a composite medium: arithmetic (λ_{ari}), harmonic (λ_{har}), geometric (λ_{geo}), square root (λ_{sqr}), Hashin-Shtrikman upper (λ^U_{HS}) and lower (λ^L_{HS}) bounds, and effective medium (λ_{eff}).

where:

$$\begin{aligned}
\lambda^U_{HS} &= \lambda_{\max} + \frac{A_{\max}}{1 - \alpha_{\max} A_{\max}},\\
\text{with: } A_{\max} &= \sum_{i=1;\ \lambda_i \neq \lambda_{\max}}^{N} \frac{n_i}{\alpha_{\max} + 1/(\lambda_i - \lambda_{\max})};\\
\lambda_{\max} &= \max(\lambda_1, \ldots, \lambda_N);\ \alpha_{\max} = \frac{1}{3\,\lambda_{\max}}\\
\lambda^L_{HS} &= \lambda_{\min} + \frac{A_{\min}}{1 - \alpha_{\min} A_{\min}},\\
\text{with: } A_{\min} &= \sum_{i=1;\ \lambda_i \neq \lambda_{\min}}^{N} \frac{n_i}{\alpha_{\min} + 1/(\lambda_i - \lambda_{\min})};\\
\lambda_{\min} &= \min(\lambda_1, \ldots, \lambda_N);\ \alpha_{\min} = \frac{1}{3\,\lambda_{\min}}.
\end{aligned} \tag{3}$$

For a two-component system consisting of pore fluid and solid rock with thermal conductivities λ_f and λ_s, respectively, Equation 3 simplifies to (Hashin and Shtrikman, 1962; Horai, 1971):

$$\lambda^U_{HS} = \lambda_s + \frac{\phi}{\frac{1}{\lambda_f - \lambda_s} + \frac{1-\phi}{3\,\lambda_s}};\ \lambda^L_{HS} = \lambda_f + \frac{1-\phi}{\frac{1}{\lambda_s - \lambda_f} + \frac{\phi}{3\,\lambda_f}}. \tag{4}$$

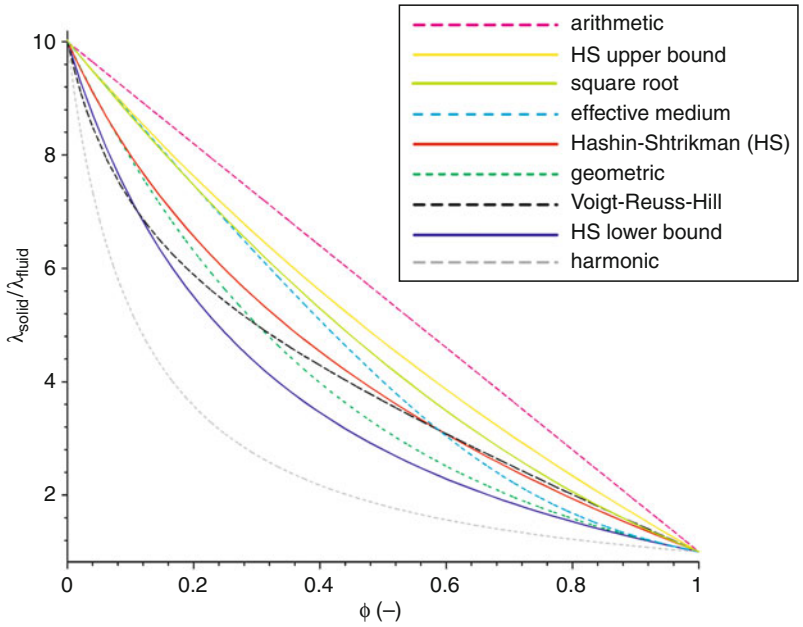
Generally, for a two-component system consisting of pore fluid and solid rock with thermal conductivities λ_f and λ_s, respectively, the implicit definition of λ_{eff} in Equation 2f can be resolved:

$$\begin{aligned}
\lambda_{eff} = \frac{1}{4}\Big\{ & 3\,\phi(\lambda_f - \lambda_s) + 2\,\lambda_s - \lambda_f \\
& + \sqrt{\begin{array}{l} 9\,\phi^2\,\lambda_s^2 + 18\,\phi\,\lambda_s\,\lambda_f - 18\,\phi^2\,\lambda_s\,\lambda_f - 12\,\phi\,\lambda_s^2 \\ + \lambda_f^2 - 6\,\phi\,\lambda_f^2 + 4\,\lambda_s\,\lambda_f + 9\,\phi^2\,\lambda_f^2 + 4\,\lambda_s^2. \end{array}} \Big\}.
\end{aligned} \tag{5}$$

Figure 3 compares the variation of thermal conductivity of a two-component system with volume fraction for these different mixing laws defined by Equations 2–5. By and large, and in particular for such a two-component system, thermal conductivity of a multiphase rock determined according to these models can be ordered as:

$$\begin{aligned}
\lambda_\perp = \lambda_{har} < \lambda^L_{HS} < \lambda_{VRH} < \lambda_{geo} < \lambda_{HS} < \lambda_{eff} \\
< \lambda_{sqr} < \lambda^U_{HS} < \lambda_{ari} = \lambda_{||}.
\end{aligned} \tag{6}$$

While only these nine models are presented and discussed here, various other mixing models are available which take into account additional factors, such as the shape of grains and voids. Several models assume spheroidal pores specifying the aspect ratio of the spheroids (Korvin, 1978, 1982; Schulz, 1981; Zimmerman, 1984, 1989; Buntebarth and Schopper, 1998; Popov et al.,



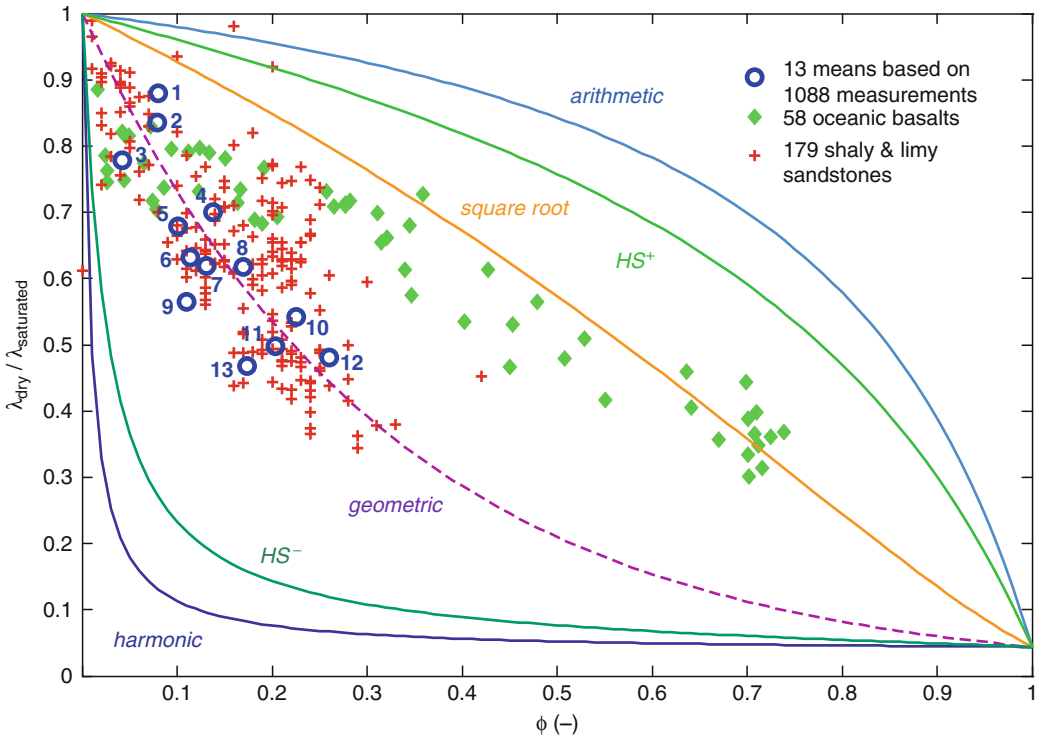
Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 3 Variation of thermal conductivity λ of a two-phase rock with porosity ϕ according to the means in Equations 2–5 for solid and fluid thermal conductivities of $\lambda_{solid} = 6\,W\,m^{-1}\,K^{-1}$ and $\lambda_{fluid} = 0.6\,W\,m^{-1}\,K^{-1}$, respectively: arithmetic (λ_{ari}); Hashin-Shtrikman upper bound (λ^{U}_{HS}); square root (λ_{sqr}); effective medium (λ_{eff}); Hashin-Shtrikman (λ_{HS}); geometric (λ_{geo}); Voigt-Reuss-Hill average (λ_{VRH}); Hashin-Shtrikman lower bound (λ^{L}_{HS}); harmonic (λ_{har}).

2003). These models require information on the geometry of the internal rock structure and differ in the way of averaging over a representative elementary volume. Horai (1991) tested the results of predictions from several different mixing models on a remarkable data set in which porosity virtually varies from 0 % to 100 %. As can be expected, most of the models tested were valid only for certain porosity ranges. Only the two-phase models of Fricke-Zimmerman (Fricke, 1924; Zimmerman, 1989) and Schulz (1981) treating pores as spheroidal inclusions in a homogeneous and isotropic material are valid over the entire range of porosity. However, they require additional information on the spheroids' aspect ratio or orientation. Given the typical ratios of rock and fluid conductivities we observe in nature, i.e. less than 10, most of the conductivity models tested work to within an accuracy of 10 %–15 %.

Based on 1 325 individual measurements on sedimentary rocks and oceanic basalts in dry and saturated condition (λ_{dry}, λ_{sat}), Figure 4 compares the variation of the ratio $\lambda_{dry}/\lambda_{sat}$ with porosity measured with that predicted by the arithmetic, upper Hashin-Shtrikman, square root, geometric, lower Hashin-Shtrikman, and harmonic mixing laws, λ_{ari}, λ^{U}_{HS}, λ_{sqr}, λ_{geo}, λ^{U}_{HS}, and λ_{har}, respectively (Equations 2–4). It is evident that no single mixing law applies equally well to all rock types. But it appears that the geometric and square root model, on average, provide the best general fits to the data. It is also clear that the upper and lower Hashin-Shtrikman means, $\lambda^{L}_{HS}$and λ^{U}_{HS}, provide much tighter bounds than the arithmetic and harmonic means, λ_{ari} and λ_{har}, respectively.

Correlations with other physical properties: Physical properties measured in well-logs can be used to infer estimates for in situ thermal conductivity. The approach is based on an extension of the mixing-model approach to the borehole scale: The volume fractions n_i of the N different mineral (or fluid) phases are either taken directly from induced gamma ray spectroscopy logs (Williams and Anderson, 1990) or determined from a joint analysis of a suitable number J of geophysical logs, such as gamma ray (GR), sonic slowness (DT, the inverse of velocity), gamma density (DEN), and neutron porosity (NPHI) (e.g., Hartmann et al., 2005; Goutorbe et al., 2006). Let $\mathbf{x}$ and $\mathbf{b}$ be vectors composed of the N volume fractions n_i and the J theoretical log responses R^j with respect to the N different phases, respectively. Then, each element R^j of vector $\mathbf{b}$ is the cumulative response of the *j*th log to all phases weighted by their volume fractions:

$$R^j = \sum_i^N n_i R_i^j, \text{ where: } \sum_i^N n_i = 1, \text{and}$$
$$\mathbf{x} = [n_1, \ldots, n_N]^T, \mathbf{b} = [R^1, \ldots, R^J]^T. \quad (7)$$



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 4 Variation of thermal conductivity ratio $\lambda_{dry}/\lambda_{sat}$ (measured in dry and saturated condition) with porosity ϕ for different rock types. Numbered open circles represent means based on measurements on 1 058 sedimentary rock samples (Kobolev et al., 1990; Popov et al., 1995, 1999a; see also Clauser, 2006): (1) 21 limestones; (2) 54 limestones; (3) 13 quartz sandstones; (4) 44 quartz silt-stones; (5) 35 conglomerates; (6) 141 quartz sandstones; (7) 33 claystones; (8) 99 polymictic sandstones; (9) 30 quartz sandstones; (10) 22 claystones; (11) 65 quartz silt-stones; (12) 99 quartz silt-stones; (13) 241 quartz silt-stones. Shown for comparison are data measured on 58 oceanic basalts (diamonds) and 179 shaly and limy sandstones (crosses). Curves labeled arithmetic, HS^+, square root, geometric, HS^-, and harmonic, correspond to the arithmetic, upper Hashin-Shtrikman, square root, geometric, lower Hashin-Shtrikman, and harmonic mixing laws, λ_{ari}, λ^{U}_{HS}, λ_{sqr}, λ_{geo}, λ^{L}_{HS}, and λ_{har}, respectively (Equations 27–29) (Clauser, 2006).

The rows of matrix $\mathbf{A}$ contain the specific responses of each log to the N rock phases:

$$\mathbf{A} = \begin{bmatrix} R_1^1 & \cdots & R_N^1 \\ \vdots & \ddots & \vdots \\ R_1^J & \cdots & R_N^J \end{bmatrix}, \tag{8}$$

and the direct and inverse problems can then be written as

$$\mathbf{Ax} = \mathbf{b} \text{ and } \mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b}, \tag{9}$$

respectively. Thus, in the direct problem, the log response vector $\mathbf{b}$ is computed from the volume fraction vector $\mathbf{x}$ and the specific log response matrix $\mathbf{A}$. In contrast, in the inverse problem, the volume fractions $\mathbf{x}$ are computed from the log responses $\mathbf{b}$ and the inverse of the specific log response matrix, $\mathbf{A}^{-1}$. Thus, for N solid rock constituents, solving the inverse requires $N - 1$ logs. Porosity does not count here, because it follows as the difference of one and the sum of the solid rock volume fractions. If more logs are available, making the problem over-determined, the inverse problem can also be solved in a least-squares sense. Once the volume fractions are known and assigned appropriate thermal conductivities, an appropriate mixing model can applied to compute rock thermal conductivity. Generally, the geometric and square root means, Equations 2d and e, often have turned out useful, but other mixing models may be appropriate in specific cases.

Assigning representative thermal conductivities to the solid rock constituents is not trivial. Tabulated values of rock thermal conductivity should be used only if they characterize specimens from the logged formations. In all other cases, these formations or their outcrops need to be sampled and these specimens tested in the laboratory. If measurements are performed at ambient conditions, the values need to be corrected for the effect of temperature, and in some cases for pressure as well.

In general, the effect of temperature is more pronounced than that of pressure. However, for greater depth

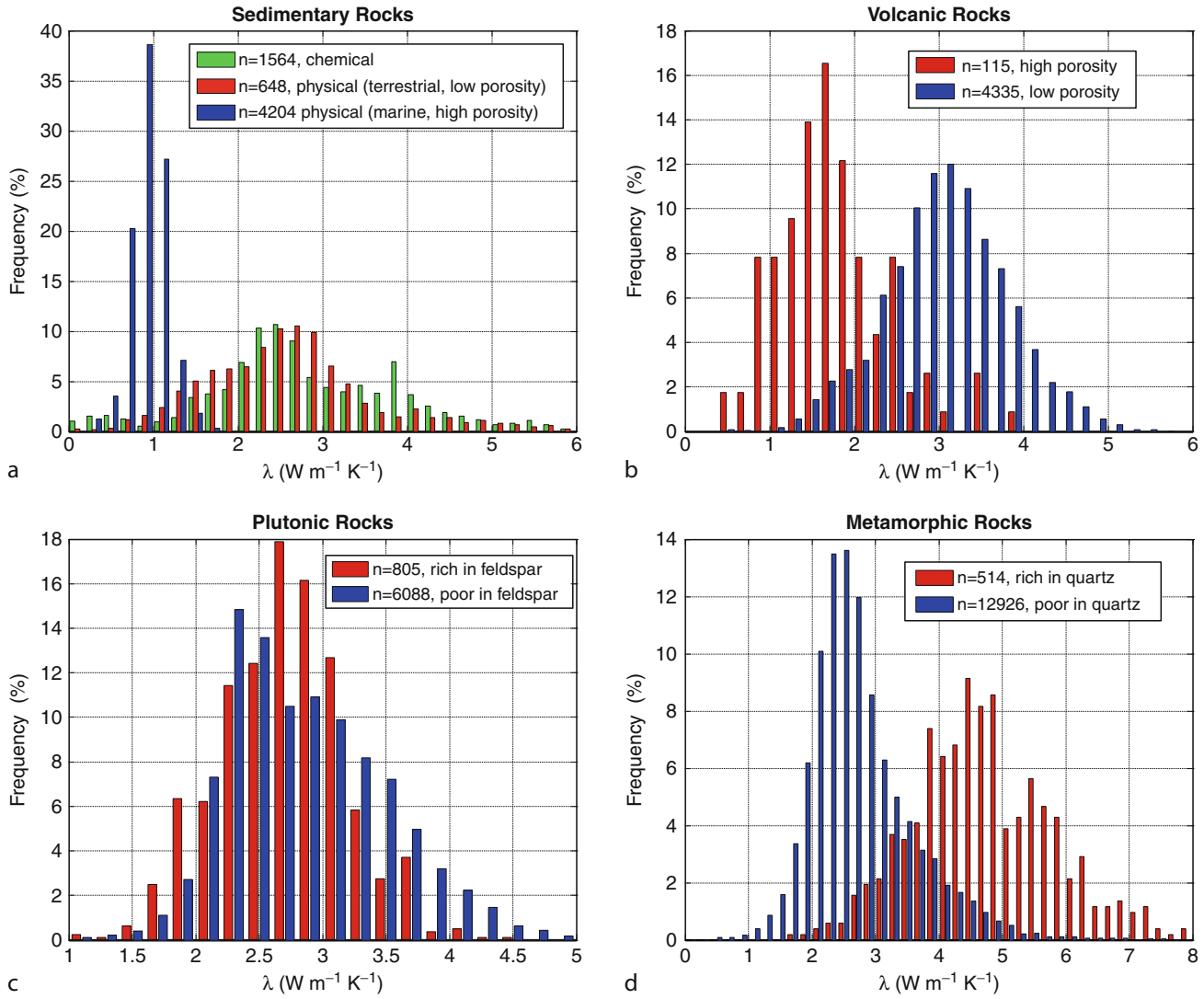
and little or less consolidated rocks it also needs to be accounted for. If commercial log interpretation software is used to perform the inversion, the theoretical log responses R_i^j with respect to the different rock constituents are usually supplied by the software. Alternatively, values for the theoretical log responses R^j can be obtained from the literature (e.g., Crain, 1986).

Thermal conductivity of minerals and rocks

Thermal conductivity of minerals is much better constrained than that of rocks, due to the well-defined crystal structure and chemical formula for each mineral. Substantial collections of mineral thermal conductivities were compiled by Birch (1942), Clark (1966), Horai and Simmons (1969), Dreyer (1974); Roy et al. (1981), Čermák and Rybach (1982), Carmichael (1984), Popov et al. (1987), Diment and Pratt (1988); Somerton (1992), Clauser and Huenges (1995), and Clauser (2006).

Rocks are less well defined. In fact, rock type is a rather poor descriptor for physical properties as it rarely characterizes the dominating factors for any given property. With respect to thermal conductivity, these comprise mineral content, porosity, pore fluid, saturation, and anisotropy for each rock type. As these factors are variable for each rock, the variation of rock thermal conductivity was characterized in a statistical manner in Clauser (2006, 2009) according to the four main diagenetic classes of rocks: sedimentary, volcanic, plutonic, and metamorphic. This discussion is summarized here.

For *sedimentary rocks* (Figure 5a), thermal conductivity is mainly controlled by porosity and sediment type.



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 5 Histograms of thermal conductivity for (**a**) sedimentary, (**b**) volcanic, (**c**) plutonic, and (**d**) metamorphic rocks (Clauser, 2009).

For *volcanic rocks* (Figure 5b), porosity is the controlling factor on thermal conductivity: Both mean and median of the high- and low-porosity histograms differ by nearly a factor of two, and the high-porosity distribution is skewed toward low conductivities. Plutonic and metamorphic rocks are generally much less porous. Here, the controlling factor is the dominant mineral phase. For *plutonic rocks* (Figure 5c), the feldspar content determines the shape of the histogram: Rocks with a high feldspar content (i.e., >60 %) display a nearly symmetrical conductivity distribution about a lower mean conductivity than rocks with low feldspar content. In spite of these differences, the means and medians for both distributions are nearly identical. For *metamorphic rocks* (Figure 5d), it is the quartz content which controls thermal conductivity: Both mean and median of the distributions for high and low quartz content differ by nearly a factor of two, similar as for *volcanic rocks* (Figure 5b) with regard to porosity. While the histogram for high quartz content rocks (mostly quartzites) is nearly symmetrical, the low quartz content histogram is strongly skewed toward low conductivities.

Radiative thermal conductivity

In addition to heat conduction, heat radiation emitted from hot bodies propagates through a sufficiently transparent and little absorbing medium as electromagnetic waves with velocity

$$c_m = c_0/n = \Lambda_m \nu_m, \tag{10}$$

where Λ_m and ν_m are radiation frequency and wavelength in the medium, and $n = c_0/c_m$ is the ratio of the speed of light in vacuum and in the medium, i.e., the real part of the index of refraction (Table 1).

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 1 Refractive index n of selected substances

Substance	$n = c_0/c_m$ (at 589 nm)
Air	1.0003
Water(H_2O)	1.33
Magnesium fluoride (MgF_2)	1.38
Calcium fluoride (CaF_2)	1.43
Rock salt (NaCl)	1.54
Quartz (SiO_2)	1.54
Glass	1.5–1.6
Carbon disulfide (CS_2)	1.63
Corundum (Al_2O_3)	1.76
Zircon ($ZrSiO_4$)	1.92
Sulfur (S)	2.00
Sinc sulfide (ZnS)	2.37
Diamond	2.42
Carborundum (SiC)	2.55
Rutile (TiO_2)	3.10
Galena (PbS) (@ 590 nm)	3.90

Radiation theory is based on Planck's distribution law defining the spectral radiance L_Λ or L_ν of a black body in the wavelength or frequency interval $d\Lambda_m$ or $d\nu_m$ in the medium in terms of the corresponding wavelength or frequency in vacuum, respectively:

$$L_\Lambda(\Lambda_m, T)\, d\Lambda_m = \frac{2hc_0^2 n^2(\Lambda, T)}{\Lambda^5 \left(e^{hc_0/(\Lambda kT)} - 1\right)} d\Lambda \quad (\mathrm{W\,m^{-2}\,sr^{-1}\,m^{-1}} = \mathrm{W\,m^{-3}\,sr^{-1}}),$$

$$L_\nu(\nu_m, T)\, d\nu_m = \frac{2h\nu^3 n^2(\nu, T)}{c_0^2 \left(e^{h\nu/(kT)} - 1\right)} d\nu \quad (\mathrm{W\,m^{-2}\,sr^{-1}\,Hz^{-1}} = \mathrm{W\,s\,m^{-2}\,sr^{-1}}), \tag{11}$$

where, $h = 6.626\,068\,96(33) \times 10^{-34}$ J s is the Planck constant, $k = 1.380\,6504(24) \times 10^{-23}$ J K^{-1} the Boltzmann constant, and $d\Lambda = n\, d\Lambda_m$. Radiance is the radiant energy flux emitted per unit solid angle in a given direction per unit projected area of a source. According to Wien's displacement law, the maximum of spectral radiance increases with temperature and decreases with wavelength (or increases with frequency) of radiation. Hofmeister (2005) estimated the variation of n(ν,T) with temperature and frequency and found both to be very small suggesting that n is approximately constant in the visible frequency range.

The intensity of radiation, I, is related to the incident intensity, I_0, radiation path, x, and opacity, ε, by $I = I_0 \exp(-\varepsilon x)$, where opacity $\varepsilon = 1/\ell$ is the reciprocal mean free path ℓ of radiation. It defines the average travel distance of photons before being absorbed or scattered. In general, opacity is a function of the radiation wavelength. Opacity in an absorbing and scattering medium comprises contributions from both of these processes:

$$\varepsilon = \alpha + \varsigma, \tag{12}$$

where α and ς are absorption and scattering coefficients, respectively. The scattering coefficient ς is usually identified with the inverse grain size, suggesting orders of magnitude ranging from 10^2 m to 10^6 m^{-1}. The absorption coefficient α may be as large as $\alpha = 7{,}000\,\mathrm{m^{-1}}$, but values below $\alpha = 1\,500\,\mathrm{m^{-1}}$ are reported for absorption coefficients in single crystal olivines with different proportions of forsterite (Mg_2SiO_4) and fayalite ($Fe_2^{2+}SiO_4$) ($Fo_{94}Fa_6$ – $Fo_{86}Fa_{14}$) in the two pass bands at 0.5 µm and 0.5 µm–6.0 µm (e.g., Shankland et al., 1979; Clauser, 1988; Hofmeister, 2005). The width and level of these pass bands depend critically on the iron content in the minerals. The complex index of refraction of all materials, m, is defined by $m = \sqrt{\varepsilon_r \mu_r} = n - i \cdot k$, where ε_r and μ_r are (complex) relative electrical permittivity and relative magnetic permeability, respectively, $n = c_0/c_m$ is the real and k the imaginary part of the index (sometimes also called extinction coefficient), and

$i = \sqrt{-1}$ is the imaginary unit. The absorption coefficient α is related to the complex index of refraction k by

$$\alpha = 4\,\pi\,k/\Lambda, \tag{13}$$

where Λ is the radiation wavelength (Aronsson et al., 1970).

Radiation therefore becomes important for rocks with a larger free mean path of radiation corresponding to smaller values of opacity or absorption coefficient and scattering coefficient. This holds in particular at larger wavelength, in the infrared part of the absorption spectrum's transmission window between about 0.5 μm–6.0 μm.

Radiated heat is diffused if photons emitted by mineral grains are scattered or absorbed by neighboring grains. If the mean free path of radiation, ℓ, is small compared to the distance to material discontinuities (such as grain boundaries) and for moderate temperature gradients, i.e., no large anisotropy in the intensity of radiation, an effective thermal conductivity

$$\lambda_{\mathrm{eff,ij}} = \lambda_{\mathrm{p,ij}} + \lambda_{\mathrm{r,ij}} \tag{14}$$

replaces λ_{ij} in Equation 1, where $\lambda_{p,ij}$ and $\lambda_{r,ij}$ are phonon and radiative thermal conductivities, respectively.

Rosseland (1930) and Clark (1957) defined the radiative thermal conductivity based on the temperature derivative of the spectral radiance of a blackbody. Following practice in engineering, Hofmeister (2005) accounts in this expression additionally for the emissivity $\eta \leq 1$ of a grainy material by applying the temperature derivative and the subsequent integration to the product of emissivity η and spectral radiance L_ν, yielding

$$\lambda_{\mathrm{r}} = \frac{4\,\pi}{3}\int_0^\infty \frac{1}{\varepsilon(\nu,T)}\,\frac{\partial(\eta(\nu,T)\,L_\nu(\nu,T))}{\partial T}\,\mathrm{d}\nu. \tag{15}$$

Different practical expressions were derived depending on which parameters can be assumed as independent of temperature or frequency (see, e.g., Clark (1957); Shankland et al. (1979); Schärmeli, 1982; Clauser, 1988; Hofmeister, 1999, 2005; Hofmeister et al., 2009). Based on grain size δ and attenuation

$$\mathrm{A} = (\ln \mathrm{I}_0 - \ln \mathrm{I}_\mathrm{t})/\delta \approx \alpha, \tag{16}$$

where α is the absorption coefficient and I_0 and I_t are incident and transmitted intensities, respectively, Hofmeister (2005) suggests the following expressions for emissivity and opacity:

$$\eta = 1 - e^{-\delta\,\alpha(\nu)} = 1 - e^{-(\ln \mathrm{I}_0 - \ln \mathrm{I}_\mathrm{t})}, \quad \varepsilon = \frac{1+\delta\,\alpha}{\delta}, \tag{17}$$

and estimates values for the product $\delta\,\alpha$ based on the interface reflectivity R (%) between two neighboring grains. Values for $\delta\,\alpha$ considered realistic for the mantle range between 5 and 10 (corresponding to a range for R between 0.07 % and 0.05 %) with a preferred value of 7 (R = 0.01 %).

The "gray body" approximation assumes opacity and emissivity as finite, constant, and independent of radiation wavelength. If the real part of the refractive index, n, and the spectral radiance, L_Λ or L_ν, are also independent of temperature and wavelength, Equation 15 simplifies into:

$$\lambda_{\mathrm{r}} = \frac{16\,\eta\,\sigma\,n^2\,T^3}{3\,\varepsilon}, \tag{18}$$

where σ = 5.670 400(40) × $10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the Stefan-Boltzmann constant. An example of magnitude is obtained when emissivity η = 0.99 and opacity ε is identified with the olivine ($Fo_{92}Fa_8$) absorption coefficient α at 1 700 K (Equation 12, neglecting contributions from scattering), with $1\,000\,\mathrm{m^{-1}} < \alpha < 1\,500\,\mathrm{m^{-1}}$. Additionally, a typical silicate value is assumed for the index of refraction, n = 1.6 (Table 1). This yields a range for radiative thermal conductivity at 1 700 K of $3.8\,\mathrm{W\,m^{-1}\,K^{-1}} > \lambda_\mathrm{r} > 2.5\,\mathrm{W\,m^{-1}\,K^{-1}}$.

Inserting Equations 16 and 17, and the temperature derivative of the spectral radiance (11) as provided by Shankland et al. (1979) into (15) yields radiative thermal conductivity as:

$$\begin{aligned}\lambda_\mathrm{r} &= \frac{4\,\pi\,\delta}{3}\int_0^\infty \frac{1-e^{-\delta\,\alpha(\nu)}}{1+\delta\,\alpha(\nu)}\,\frac{\partial(L_\nu(\nu,T))}{\partial T}\,\mathrm{d}\nu \\ &= \frac{8\,\pi\,\delta\,\mathrm{n\,T^3\,k^4}}{3\,c_0^2\,h^3}\sum\int_{lower}^{upper}\frac{1-e^{-\delta\,\alpha(\nu)}}{1+\delta\,\alpha(\nu)} \\ &\times\left[\frac{n\,e^x\,x^4+}{e^x-1}+\frac{2\,T\,x^3}{e^x-1}\,\frac{\partial n}{\partial T}\right]dx\ (\mathrm{W\,m^{-1}\,K^{-1}}),\end{aligned} \tag{19}$$

where x = hν/(k T) and dx = h dν/(k T), and the summation allows for the transparent regions above and below the strong absorption bands in the visible part of the spectrum. The second term of the sum in the integrand vanishes if $\partial n/\partial T \approx 0$ as suggested by Hofmeister (2005). Because α varies nonlinearly with frequency or wavelength, and the cutoff frequencies in the integral in (19) depend on δ and α, Hofmeister (2005) evaluated the integral numerically for polynomials in T whose exponents ranged from 0 to 6, thus yielding other than a purely cubic relationship between radiative thermal conductivity and temperature.

Variation with temperature

Since the pioneering experiments of Eucken (1911), thermal conductivity of minerals and rocks is known to decrease with temperature, generally with its inverse. Eucken's empirical result was corroborated theoretically by Peierls (1929) based on Debye's (1914) theory of phonon scattering. Eucken (1911) observed in his experiments

in contrast to crystals an increase in thermal conductivity of amorphous siliceous glass, a clear indication of radiative heat transfer.

The decrease is primarily due to the decrease of phonon (or lattice) thermal conductivity λ_p with temperature and to a smaller degree to thermal cracking. Since the thermal expansion coefficient increases with temperature (but differently for all minerals) differential expansion may create contact resistances between mineral grains. The effect of contact resistance is less pronounced in water-saturated than in dry rocks, the condition in which most rocks are tested at elevated temperatures. For single-mineral aggregates, a linear relationship between temperature and thermal resistivity, λ^{-1}, discriminates between contributions which depend on temperature T and others which do not, such as micro-cracks, grain boundaries, shape, and orientation of crystals and their fragments:

$$\lambda^{-1}(T) = c_1 + c_2\ T, \tag{20}$$

where λ is in $W\,m^{-1}\,K^{-1}$ and T is in K. By measuring thermal conductivity λ and plotting its inverse, thermal resistivity, λ^{-1}, versus temperature, constants c_1 and c_2 are obtained from intercept and slope of a linear regression. Table 2 provides values for the constants c_1 and c_2 in Equation 20 which may be used to infer the temperature dependence of thermal resistivity for some single-mineral aggregates (Clark, 1969).

Based on measurements on 113 samples of metamorphic rocks from the KTB research borehole in Germany in the temperature range 50 °C–200 °C, Buntebarth (1991) determined mean values for the constants c_1 and c_2 in Equation 20 for gneissic and metabasitic rocks. The arithmetic means determined from measurements on 66 gneiss samples are $\bar{c}_1 = 0.16(3)\,W^{-1}\,m\,K$ and $\bar{c}_2 = 0.37(14) \times 10^{-3}\,W^{-1}\,m$. The corresponding means determined from measurements on 36 metabasite samples are $\bar{c}_1 = 0.33(3)\,W^{-1}\,m\,K$ and $\bar{c}_2 = 0.22(14) \times 10^{-3}\,W^{-1}\,m$.

In contrast to phonon conductivity λ_p, the radiative contribution to thermal conductivity, λ_r, generally increases with the cube of temperature (see above). Thus, measurements of thermal conductivity as function of temperature generally first exhibit a decrease with temperature until, from about 1 000 °C–1 200 °C onward, the radiative component balances and sometimes even reverses the decreasing trend.

The temperature dependence of rock thermal conductivity was characterized in a statistical manner in Clauser (2006, 2009) according to the four main diagenetic classes of rocks: sedimentary, volcanic, plutonic, and metamorphic. This discussion is summarized here (Figure 6).

For *sedimentary rocks* (Figure 6a) up to 300 °C there is a reduction by nearly a factor of two, both for physical and chemical sediments. Above 300 °C, the decrease in thermal conductivity is less, but it is stronger for chemical sediments than for physical sediments. However, there are very few data for this temperature range, which makes this last observation statistically weak. Above 300 °C, the mean thermal conductivity of sediments varies between $1.0\,W\,m^{-1}\,K^{-1}$ – $1.5\,W\,m^{-1}\,K^{-1}$.

Volcanic rocks (Figure 6b) vary quite differently with temperature depending on their opacity, i.e., on how well they transmit thermal energy by radiation. Due to this additional "radiative thermal conductivity," volcanic glasses and rocks with small iron content experience an increase in thermal conductivity for temperatures above 800 °C–1 000 °C (e.g., Clauser, 1988; Hofmeister et al., 2009). In contrast, thermal conductivity of conduction dominated rocks, such as rocks with high iron content, decreases with temperature. An inversion of this trend is indicated by few available high-temperature measurements (above 1 300 °C) but with too few measurements for a statistical appraisal. At about 1 000 °C thermal conductivity for these rocks is at about 50 % of the room-temperature value. Again, there are few data points above 700 °C.

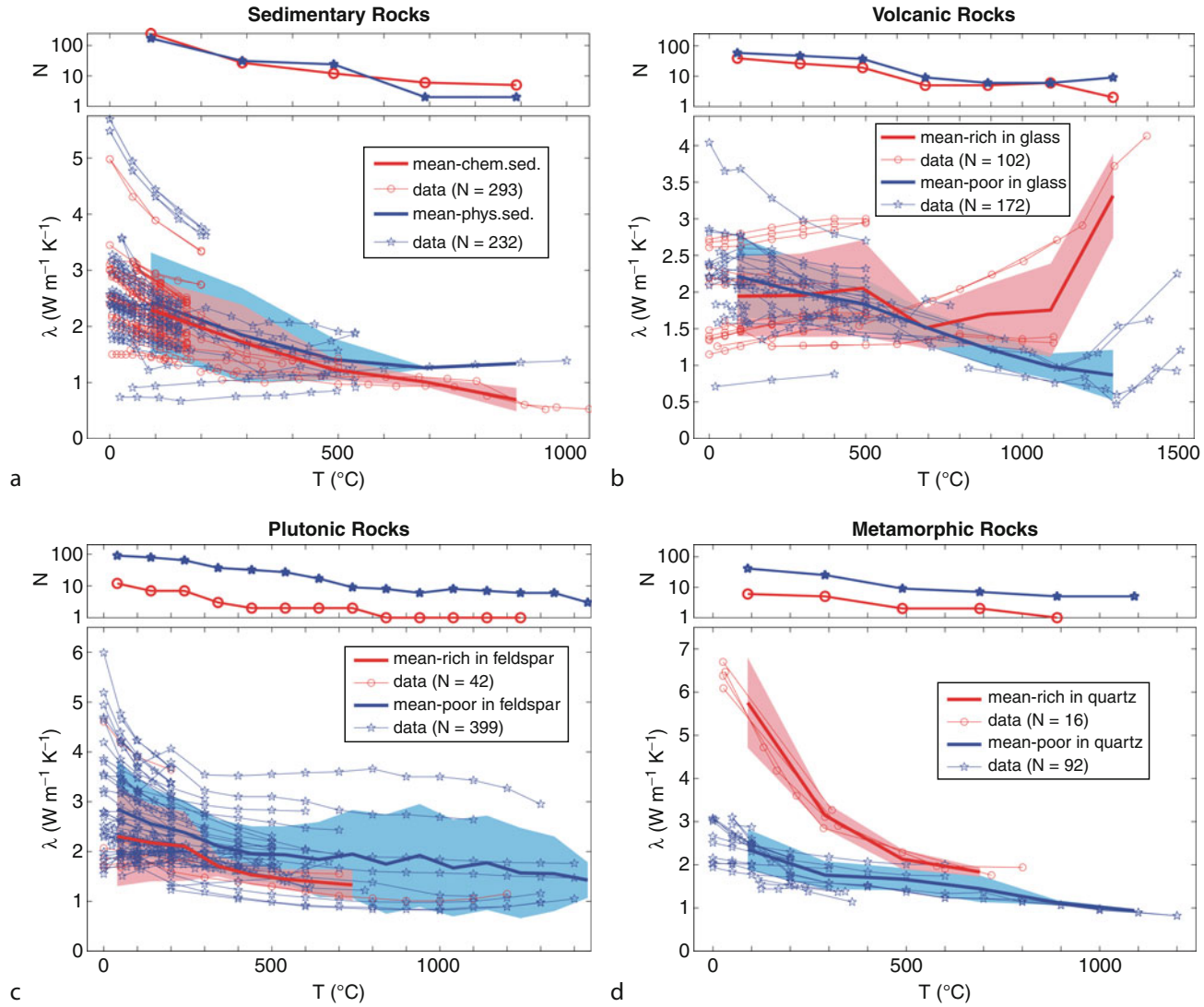
P*lutonic rocks* (Figure 6c) show no strong radiative contribution. At temperatures above 600 °C, thermal conductivity decreases only very little. However, in these rocks the variation of thermal conductivity with temperature depends on their feldspar content. For rocks enriched in feldspar, thermal conductivity decreases little up to 300 °C, while for those poor in feldspar the decrease is stronger, becoming more gentle above 300 °C, and spreading an additional 20 % over the next 1 000 K. The different behavior of rocks with high feldspar content is due to the increase in thermal conductivity with temperature of some plagioclase feldspars (e.g., Höfer and Schilling, 2002; Petrunin et al., 2004) which compensates the decrease in thermal conductivity with temperature observed for most other minerals and rocks. Other notable exceptions are fused silica as well as volcanic and silica glasses.

For *metamorphic rocks* (Figure 6d), the decrease of thermal conductivity with temperature depends on the content in a dominant mineral phase, similar as for plutonic rocks. For quartzites, the decrease is strong, by nearly a factor of three up to a temperature of about 500 °C with only a very mild further decrease beyond this

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 2 Values of c_1 and c_2 in Equation 20 for single-mineral aggregates; data: Clark (1969)

Mineral	T (°C)	$c_1 \times 10^3$ ($W^{-1}\,m\,K$)	$c_2 \times 10^3$ ($W^{-1}\,m$)
Halite, NaCl	0–400	−52.55	0.788
Periclase, MgO	100–800	−21.50	0.127
Corundum, Al_2O_3	100–800	−28.66	0.155
Quartz, SiO_2[a]	100–400	62.10	0.387
Spinel, $MgAl_2O_4$	100–1 000	19.11	0.122
Zircon, $ZrSiO_4$	100–800	131.37	0.093
Forsterite, Mg_2SiO_4,	100–600	85.98	0.282
Enstatite, ferrosilite, $(Mg_2,Fe_2)SiO_3$	100–300	200.63	0.222

[a]Single SiO_2 crystal, heat flowing ⊥ to optical axis



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 6 Variation of thermal conductivity with temperature for (**a**) sedimentary, (**b**) volcanic, (**c**) plutonic, and (**d**) metamorphic rocks. Color shading indicates a range defined by plus and minus one standard deviation and N is the number of data at each temperature (Clauser, 2009).

temperature. For rocks poor in quartz the decrease in conductivity is not quite as strong, amounting to about one third of the room-temperature value up to 300 °C. Then it remains roughly constant up to 500 °C and decreases again to about one third of the room-temperature value up to 750 °C.

In summary, for moderate temperatures thermal conductivity of rocks is well described by a linear relationship with inverse temperature, similar as in Equation 20. For this temperature range several approaches are available for inferring thermal conductivity at elevated temperatures. Based on the analysis of available tabulated data of thermal conductivity as function of temperature Zoth and Hänel (1988) suggested the following form:

$$\lambda(T) = A + \frac{B}{350 + T} (0\,^{\circ}C \le T \le 800\,^{\circ}C), \qquad (21)$$

where average values of the coefficients A and B for different rock types are given in Table 3.

Linear relationships between temperature and thermal resistivity, such as Equations 20 and 21, discriminate between temperature-dependent contributions and other factors, which are independent of temperature, such as micro-cracks, grain boundaries, pore volume, mineralogical composition, shape, and orientation of crystals and their fragments.

Sass et al. (1992) and Vosteen and Schellschmidt (2003) distinguish between the effects of composition

and temperature. They propose a general empirical relation for $\lambda(T)$, the thermal conductivity in W m^{-1} K^{-1} at temperature T in °C, as a function of λ_0, the thermal conductivity at 0 °C:

$$\lambda(T) = \frac{\lambda_0}{a + T\left(b - \frac{c}{\lambda_0}\right)} \text{ or } \frac{\lambda_0}{\lambda(T)} = \underbrace{a}_{\text{intercept}} + \underbrace{(b - \frac{c}{\lambda_0})}_{\text{slope}}\ T \quad (22)$$

For different rock types, the slopes and intercepts of this equation can be determined from linear regressions of Equation 22 yielding a mean intercept ā and its uncertainty Δa. Coefficients b and c and associated uncertainties σ_b and σ_c are determined from a second linear regression of the different slopes ($b - c/\lambda_0$) as a function of $1/\lambda_0$ (Table 4).

Since thermal conductivity is usually measured at room temperature, λ_0 is expressed as a function of λ_{25}, the room temperature thermal conductivity, by Sass et al. (1992) for crystalline rocks (felsic gneiss to amphibolite) as:

$$\lambda_0 = \lambda_{25}\left(1.007 + 25\left(0.0037 - \frac{0.0074}{\lambda_{25}}\right)\right). \quad (23)$$

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 3 Values for constants A and B in Equation 21 for different rock types (Zoth and Hänel, 1988)

Rock type	T (°C)	A (W m^{-1} K^{-1})	B (W m^{-1})
1. Rock salt	−20 to 0	−2.11	2 960
2. Limestones	0–500	0.13	1 073
3. Metamorphic rocks	0–1 200	0.75	705
4. Acidic rocks	0–1 400	0.64	807
5. Basic rocks	50–1 100	1.18	474
6. Ultra-basic rocks	20–1 400	0.73	1 293
7. Rock types (2)–(5)	0–800	0.70	770

Vosteen and Schellschmidt (2003) find for magmatic and metamorphic rocks:

$$\lambda_0 = 0.53\ \lambda_{25} + 0.5\ \sqrt{1.13\ \lambda_{25}^2 - 0.42\ \lambda_{25}}, \quad (24)$$

and for sedimentary rocks:

$$\lambda_0 = 0.54\ \lambda_{25} + 0.5\ \sqrt{1.16\ \lambda_{25}^2 - 0.39\ \lambda_{25}}. \quad (25)$$

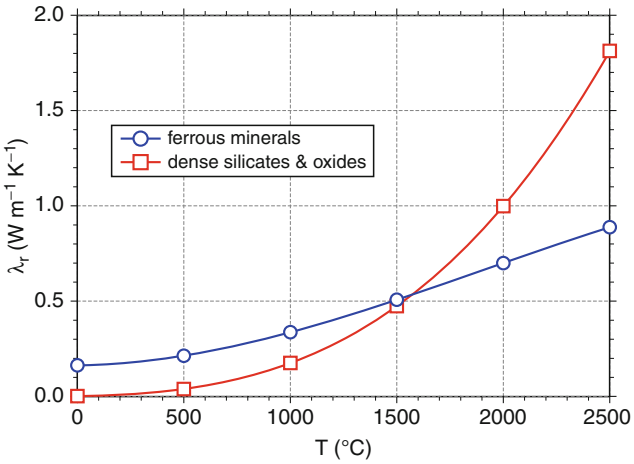
Hofmeister (1999) provided a detailed analysis of heat transport based on an analysis of phonon lifetimes obtained from infrared reflectivity. It accounts for the variation of the phonon contribution λ_p to thermal conductivity with both temperature and pressure as well as for the pressure dependent radiative contribution λ_r. It allows approximation of thermal conductivity for mantle conditions if: (1) K'_0, the pressure derivative of the isothermal bulk modulus K_T, is constant: $K'_0 = dK_T/dP$ = const; (2) the variations of the bulk modulus with temperature and pressure are mutually independent; (3) the pressure derivative of the thermodynamic Grüneisen parameter γ (Equation 12 or 13 in *Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat*, this volume) is constant: $d\gamma/dP = f$. For mantle material, γ varies from 1 to1.4, K'_0 from 4 to 5, and $f \approx 0$ vanishes approximately. According to Hofmeister (1999) and within the uncertainty of these parameters, thermal conductivity in the mantle is:

$$\lambda(T,P) = \lambda_{298\,K,\,101.33\,kPa} \times (298/T)^a (1 + K'_0 P/K_T) e^{-(4\gamma+1/3)\int_{298}^{T}\alpha(T')dT'} + \lambda_r, \quad (26)$$

where T is absolute temperature, $\lambda_{298\ K,\ 101.33\ kPa}$ is thermal conductivity at room temperature and atmospheric pressure, α(T) is volume coefficient of thermal expansion as a function of temperature, and the exponent a is the fitting parameter. The radiative contribution λ_r in (26) may be approximated by Equations 15, 18, or 19. Hofmeister (1999) provides alternative expressions for λ_r for ferrous minerals or dense silicates and oxides (Figure 7):

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 4 Coefficients a, b, and c in Equation 22 and associated uncertainties Δa and σ_b, σ_c; Δa is the error of the mean intercept ā for all rock types of the linear regressions of the normalized thermal resistance $\lambda_0/\lambda(T)$ as a function of temperature T; σ_b and σ_c are the errors defined by the linear regression of the slopes ($b - c/\lambda_0$) as a function of the thermal resistance $1/\lambda_0$ (see Equation 22)

Rock type	ā (–)	Δa (%)	$b \times 10^3$ (K^{-1})	$\sigma_b \times 10^3$ (K^{-1})	$c \times 10^3$ (W m^{-1} K^{-2})	$\sigma_c \times 10^3$ (W m^{-1} K^{-2})	T (°C)	Reference
Basement rocks I (from felsic gneiss to amphibolite)	1.007	–	3.6	–	7.2	–	0–250	Sass et al., 1992
Basement rocks II (magmatic and metamorphic)	0.99	1	3.0	1.5	4.2	0.6	0–500	Vosteen and Schellschmidt, 2003
Sediments	0.99	1	3.4	0.6	3.9	1.4	0–300	Vosteen and Schellschmidt, 2003



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 7 Variation of radiative thermal conductivity λ_r of ferrous minerals, dense silicates, and oxides with temperature according to Equation 27 (Hofmeister, 1999).

ferrous minerals :

$$\lambda_r = 0.01753 - 1.0365 \times 10^{-4} T + 2.2451 \times 10^{-7} T^2 - 3.407 \times 10^{-11} T^3;$$

dense silicates and oxides

$$\lambda_r = 8.5 \times 10^{-11} T^3 \ (\lambda_r \text{ in W m}^{-1}\text{ K}^{-1} \text{ and T in K}). \quad (27)$$

Figure 1 in the companion article in this volume (see *Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat*) shows the variation of phonon thermal conductivity λ_p with temperature in an average crust of density $\rho = 2\,700\,\text{kg m}^{-3}$ molar mass of $0.22178\,\text{kg mol}^{-1}$. It is derived from $\lambda = \kappa\,\rho\,c$ as the product of thermal capacity $\rho\,c_P$ and thermal diffusivity κ calculated according to Equation 20 (in Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat, this volume"; Whittington et al., 2009). Due to the balancing increase of specific heat capacity with temperature, the decrease in thermal conductivity with temperature is less than in thermal diffusivity by a factor of about 1.3. Assuming a constant density throughout the crust, this implies that the increase and decrease in density due to the increase in pressure and temperature, respectively, partly cancel each other and that these changes are small compared to those of specific heat capacity and thermal diffusivity.

For mantle minerals, such as olivine and its high-pressure polymorphs, the β- and γ-spinels wadsleyite and ringwoodite, respectively, Xu et al. (2004) fitted phonon thermal conductivity measured to 1 373 K and 20 GPa to an exponential equation in temperature yielding values for the exponent between −0.406(35) and −0.537(11) suggesting that fitted phonon thermal conductivity varies with $T^{-1/2}$ (Table 5).

Variation with pressure

The effect of pressure on phonon thermal conductivity λ_p is different in two distinct pressure ranges. First, fractures and micro-cracks (developed during stress release after sampling) begin to close with increasing pressure. This reduces thermal contact resistance as well as porosity which is usually filled with a low conductivity fluid. This process ends when a pressure of about 15 MPa is reached. A compilation of measurements on various sedimentary, volcanic, plutonic, and metamorphic rocks (Clauser and Huenges, 1995) indicates that this effect accounts for an increase of about 20 % relative to thermal conductivity at atmospheric pressure. A further pressure increase to 40 MPa does not affect thermal conductivity significantly. If pressure is increased further, however, a second process becomes effective, the reduction of intrinsic porosity, i.e., voids which are not created by stress release. For granite and for metamorphic rocks, data indicate an increase of thermal conductivity by about 10 % within the pressure range 50 MPa–500 MPa.

For mantle minerals, such as olivine and its high-pressure polymorphs, the β- and γ-spinels wadsleyite and ringwoodite, respectively, Xu et al. (2004) determined values for the pressure coefficient between $0.022\,\text{GPa}^{-1}$ and $0.032\,\text{GPa}^{-1}$ for phonon thermal conductivity measured to 1 373 K and 20 GPa (Table 5). A table of numerical values for pressure derivatives of phonon thermal conductivity measured by a variety of authors was compiled by Hofmeister et al. (2009). Most values for $\lambda^{-1}(\partial\lambda/\partial P)$ fall into the range $0.04\,\text{GPa}^{-1}$–$0.36\,\text{GPa}^{-1}$, exceeded only by values of $0.69\,\text{GPa}^{-1}$ and $0.5\,\text{GPa}^{-1}$ for sulfur and quartz$_{\perp c}$ (measured perpendicular to the optical *c*-axis), respectively. Osako et al. (2004) fitted thermal conductivity measured to 1 100 K and 8.3 GPa on isotropic, single-crystal garnet and anisotropic olivine ($Fo_{93}Fa_7$; in three crystallographic directions) as upper and lower mantle constituents, respectively, to a linear equation in pressure (Table 6). A pressure dependence of garnet and olivine was found on the order of $4\,\%\,\text{GPa}^{-1}$–$5\,\%\,\text{GPa}^{-1}$ and $3\,\%\,\text{GPa}^{-1}$–$4\,\%\,\text{GPa}^{-1}$, respectively.

In contrast, radiative thermal conductivity λ_r was found much less variable with pressure than with temperature (Clark, 1957). In particular, this holds once the spectral radiance overlaps the infrared pass band in the absorption spectrum, at temperatures above 1 900 K (Hofmeister, 2005).

Variation with other factors

Apart from temperature and pressure, thermal conductivity also varies with porosity, pore fluid, saturation, dominant

mineral phase, and anisotropy. These effects are summarized here from the detailed discussion in Clauser (2006):

For large *porosity* (i.e., $\phi >> 1\,\%$) thermal conductivity of the saturating fluid affects significantly the bulk rock thermal conductivity. The influence varies with the thermal conductivity of the saturants, e.g., water, oil, natural gas, air. The resulting bulk thermal conductivity can be estimated from a suitable mixing model, e.g., Equations 2–4.

The effect of partial *saturation* is different for porous or fractured rocks. In porous rocks, porosity comprises both bulk pore space and bottlenecks formed by the contact between individual grains. Dry bottlenecks act as thermal contact resistances between grains, while the bulk pore volume contributes proportionally to the effective rock thermal conductivity. In fractured rocks, in contrast, there are no bottlenecks between grains as in porous rocks, and the small void volume in the fractures corresponds to the bulk pores space of porous rocks.

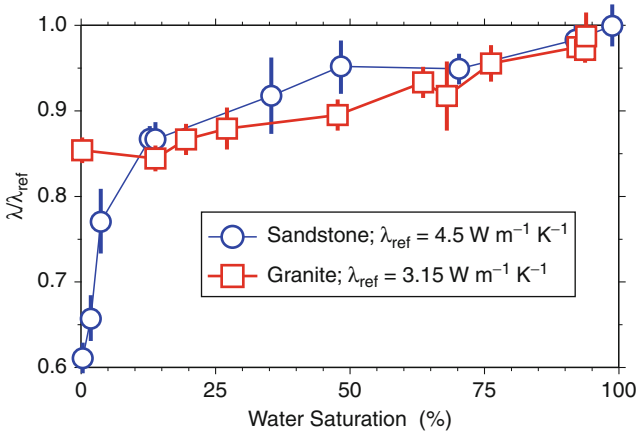
Figure 8 illustrates these two effects for a water-saturated medium-porosity sandstone and a low-porosity granite: Starting completely dry with an unsaturated conductivity of about 60 % of the saturated value, a level of 85 % is reached for the sandstone at about 10 % saturation. The 15 % conductivity residual is then spread almost linearly over the remaining 90 % of saturation. Physically this observation indicates that the filling of inter-granular bottlenecks accounting for only about 10 %−15 % of the total porosity significantly reduces the contact resistances between the individual grains. In contrast, the replacement of low conductivity air by the more conductive fluid in the major part of the pore volume accounts for the second effect. If only fractures contribute to the total porosity, as in the granite, there are no bottlenecks and only the second effect is observed: Starting completely dry with an unsaturated conductivity of about 85 % of the saturated conductivity, a quasi linear increase is observed

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 5 Reference values λ_{298} and κ_{298} at 298 K, pressure coefficients a and a′, and fitting functions for thermal conductivity λ and thermal diffusivity κ of lower mantle minerals with absolute temperature T and pressure P (Xu et al., 2004)

		$\lambda = \lambda_{298}\,(298/T)^{1/2}\,(1 + a\,P)$		$\kappa = \kappa_{298}\,(298/T)^{n}\,(1 + a'P)$		
Mineral	P (GPa)	λ_{298} (W m^{-1} K^{-1})	a (GPa^{-1})	$\kappa_{298} \times 10^6$ (m^2 s^{-1})	n	a′ (GPa^{-1})
Olivine	4–10	4.13(11)	0.032(3)	1.31(5)	0.681(22)	0.036(4)
	4	4.49(4)	–	1.29(5)	0.563(35)	–
	7	5.19(4)	–	1.74(4)	0.720(26)	–
	10	5.56(4)	–	1.84(5)	0.723(31)	–
Wadsleyite	14	8.10(4)	0.023	2.55(3)	0.721(13)	–
Ringwoodite	20	9.54(5)	0.022	3.09(4)	0.793(17)	–

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 6 Coefficients and fitting functions for the variation thermal conductivity λ and thermal diffusivity κ of garnet and olivine with absolute temperature T and pressure P (Osako et al., 2004)

Garnet				Olivine				
$\lambda = C_0 + C_1/T$		$\lambda = A_0 + A_1\,P$			$\lambda = C_0 + C_1/T$		$\lambda = B_0\,\exp(B_1\,P)$	
C_0 (W m^{-1} K^{-1})	C_1 (W m^{-1})	A_0 (W m^{-1} K^{-1})	A_1 (W m^{-1} K^{-1} GPa^{-1})		C_0 (W m^{-1} K^{-1})	C_1 (W m^{-1})	B_0 (W m^{-1} K^{-1})	B_1 (GPa^{-1})
2.01(8)	704(43)	3.48(33)	0.160(26)	[100]	1.91(28)	2,088(163)	6.61(13)	0.038(5)
				[010]	0.84(36)	1,377(157)	3.98(15)	0.042(5)
				[001]	2.08(38)	1,731(86)	5.91(25)	0.034(5)
$\kappa = c_0 + c_1/T$		$\kappa = a_0 + a_1\,P$			$\kappa = c_0 + c_1/T$		$\kappa = b_0\,\exp(b_1\,P)$	
$c_0 \times 10^6$ (m^2 s^{-1})	$c_1 \times 10^6$ (m^2 s^{-1} K)	$a_0 \times 10^6$ (m^2 s^{-1})	$a_1 \times 10^6$ (m^2 s^{-1} GPa^{-1})		$c_0 \times 10^6$ (m^2 s^{-1})	$c_1 \times 10^6$ (m^2 s^{-1} K)	$b_0 \times 10^6$ (m^2 s^{-1})	b_1 (GPa^{-1})
0.29(6)	374(31)	1.19(6)	0.046(1)	[100]	−0.06(11)	938(46)	2.50(4)	0.033(5)
				[010]	−0.13(8)	626(45)	1.52(6)	0.040(7)
				[001]	−0.03(17)	832(98)	2.16(14)	0.035(3)



Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Figure 8 Variation of thermal conductivity with partial saturation for a sandstone (circles; $\phi = 18\,\%$) and granite (squares; $\phi = 1\,\%$) saturated with water and standard deviations (bars); values normalized by reference thermal conductivities shown in legend (Clauser, 2006 based on data of Reibelt, 1991).

due to the replacement of low conductivity air until the 100 % level is reached for complete saturation.

Anisotropy of sedimentary and metamorphic rocks is due to the conditions of their formation. Anisotropy exists on several scales: (1) On the microscopic scale, many minerals are anisotropic; (2) on the laboratory scale, thermal conductivity of many rocks is also anisotropic. However, even if rocks are composed of anisotropic minerals, random orientation of the crystals within the rock may render the rock's bulk thermal conductivity isotropic on a macroscopic scale; (3) on a still larger scale, if rocks are exposed to folding, orogenic or other tectonic processes, thermal conductivity of the resulting rock formation may be anisotropic.

As a result, thermal conductivity parallel to the direction of layering or foliation, $\lambda_{||}$, is greater than thermal conductivity in the perpendicular direction, $\lambda_{\perp}$. The factor of anisotropy, the ratio $\lambda_{||}/\lambda_{\perp}$, generally falls into the range 0.9–3, with most values between 1 and 2 (e.g., Clauser and Huenges, 1995; Popov and Mandel, 1998; Popov et al., 1999a, b; Clauser, 2006; Davis et al., 2007). For sedimentary rocks a general trend has been reported of decreasing $\lambda_{\perp}$ with factor of anisotropy $\lambda_{||}/\lambda_{\perp}$, but no such trend was identified for metamorphic rocks (Clauser, 2006).

Thermal diffusivity

Thermal diffusivity is required in the analysis of transient heat transfer. A sizable compilation of room-temperature data of phonon thermal diffusivity κ_p measured on various mantle minerals is given by Hofmeister et al. (2009) from measurements by various groups.

If both conductivity and thermal capacity are known, thermal diffusivity κ can be calculated from $\kappa = \lambda/(\rho\, c)$. As for steady-state thermal conduction, transient heat diffusion in most of the Earth's crust and mantle is caused by scattering of quantized lattice vibrations, the phonons, and by diffusive (as opposed to ballistic) radiation of photons. These two processes are described by phonon thermal conductivity λ_p and radiative thermal conductivity, λ_r, respectively, the sum of which is often termed effective thermal conductivity, λ_{eff}. As thermal diffusivity is the ratio of thermal conductivity and thermal capacity, it is also influenced by the variation of density and specific heat capacity. This is of particular interest with respect to the variation with temperature.

Measuring techniques

All of the transient laboratory methods used to determine thermal conductivity are useful to determine thermal diffusivity as well. Recently, heat pulse (Schilling, 1999; Höfer and Schilling, 2002; Gibert et al., 2003) and laser flash methods (Parker et al., 1961; Blumm and Lemarchand, 2002; Hofmeister, 2006) have been used for measurements at high temperature. Compared to other methods their advantage lies in a reduction or even complete absence of physical contacts between samples on the one hand and temperature sensors and heat sources on the other hand. While most methods measure the effective diffusivity comprising contributions from phonon conduction and diffused heat radiation, the laser flash method yields the diffusive component without contributions from heat radiation. This difference becomes important particularly at high temperatures (for a critical assessment see e.g., Hofmeister et al., 2009).

Variation with temperature

Thermal diffusivity κ of rocks varies even more strongly with temperature than thermal conductivity λ. This is caused by the opposite behavior of thermal conductivity and thermal capacity (ρ c) with respect to temperature. Because of several self-compensating factors, thermal capacity (ρ c) with few exceptions generally varies within $\pm 20\,\%$ of $2.3\,\mathrm{MJ\,m^{-3}\,K^{-1}}$ for the great majority of minerals and rocks (Beck, 1988). This is confirmed by a linear regression of thermal diffusivity on thermal conductivity which was measured on a suite of metasedimentary, volcanic, magmatic, and metamorphic rocks together with density and specific heat capacity (Mottaghy et al., 2005):

$$\kappa = \frac{\lambda}{\rho\, c} = \frac{\lambda}{2.3} = 0.44\,\lambda\,(\kappa \text{ in } 10^{-6}\,\mathrm{m^2\,s^{-1}}). \qquad (28)$$

A linear regression of thermal capacity as a function of temperature yields also a linear relationship. This allows to determine thermal diffusivity $\kappa(T)$ at any temperature,

based only on the known variation of thermal conductivity λ(T) with temperature (Mottaghy et al., 2005):

$$\kappa(T) = \frac{\lambda(T)}{2.134 + 0.0044\,T} \quad (\kappa \text{ in } 10^{-6}\,\mathrm{m^2\,s^{-1}};\ \mathrm{T\ in\ ^\circ C}). \tag{29}$$

Thus, thermal diffusivity can be derived from thermal conductivity and vice versa. For the suite of rocks studied by Mottaghy et al. (2005), thermal conductivity decreased by 4 %–7 % in the range 1–100 °C while thermal diffusivity decreased by 18 %–22 %.

Thermal diffusivity was measured at temperatures of up to 550 °C by Ray et al. (2006) on a suite of 16 samples comprising Archean granulitic rocks which are considered a major component of the middle and lower continental crust. Phonon scattering was found dominating heat transport with radiative diffusion of photons setting in at 450 °C for most, but not all rocks. Based on their measurements, Ray et al. (2006) propose an equation by which thermal diffusivity κ(T) at elevated temperature below 450 °C can be derived from room temperature values, κ_{rt}:

$$\kappa(T) = 0.7 + 144\frac{\kappa_{rt} - 0.7}{T - 150} \quad (\kappa \text{ in } 10^{-6}\,\mathrm{m^2\,s^{-1}};\ \mathrm{T\ in\ K}), \tag{30}$$

They also proposed an additional term in Equation 30, proportional to T^3, by which the radiative contribution above 450 °C is fitted. However, this assumes a similar variation with temperature of thermal conductivity and thermal diffusivity and neglects the additional variation of specific heat capacity. This, in fact, makes thermal diffusivity vary stronger with temperature than thermal conductivity, which is why the T^3-term is omitted here. Whittington et al. (2009) measured the phonon component κ_p of thermal diffusivity at temperatures of up to 1 260 K on garnet schist, leucogranite, and welded rhyolitic ash-flow tuff using laser flash analysis. This characterizes purely the phonon heat transfer component without any radiative contribution, in contrast to the values discussed above. Below and above the transition between α- and β-quartz at 846 K (~573 °C) the data are reasonably fitted by

$$\kappa_p(T) = \begin{cases} -0.062 + \frac{567.3}{T}; & \mathrm{T} > 846\ \mathrm{K} \\ 0.732 - 0.000135\,T; & \mathrm{T} < 846\ \mathrm{K} \end{cases} \quad (\kappa \text{ in } 10^{-6}\,\mathrm{m^2\,s^{-1}};\ \mathrm{T\ in\ K}), \tag{31}$$

assuming an average molar mass of 0.22178 kg mol^{-1} and an average density of 2 700 kg m^{-3} for the crust (see Figure 1 in *Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat*, this volume). The leucogranite and rhyolite samples were homogeneous and isotropic. The schist was anisotropic owing to alternating mica- and quartz-rich layers, requiring testing in the direction parallel and perpendicular to foliation.

For mantle minerals, such as olivine and its high-pressure polymorphs, the β- and γ-spinels wadsleyite and ringwoodite, respectively, Xu et al. (2004) fitted phonon thermal diffusivity measured to 1 373 K and 20 GPa to an exponential equation in temperature yielding values for the exponent between −0.563(35) and −0.793(17) suggesting that fitted phonon thermal conductivity varies with $T^{-½}$–T^{-1} (Table 5).

Pertermann and Hofmeister (2006) measured thermal diffusivity on oriented single crystals and polycrystalline samples of olivine-group minerals with the laser-flash method at temperatures of up to about 1 500 °C. They fitted the data to a second order polynomial in T:

$$\kappa_p(T) = a + b/T + c/T^2 (\mathrm{T\ in\ K}). \tag{32}$$

Values for the coefficients a, b, and c fitted to data measured on single crystal and polycrystalline samples are shown in Table 7.

Osako et al. (2004) fitted thermal diffusivity measured to 1 100 K and 8.3 GPa on isotropic, single-crystal garnet and anisotropic olivine ($Fo_{93}Fa_7$; in three crystallographic directions) as upper and lower mantle constituents, respectively, to a linear equation in inverse temperature (Table 6). They found a strong anisotropy in olivine which they assume to prevail throughout the olivine stability field in the mantle down to 410 km.

Variation with pressure

Tommasi et al. (2001) measured thermal diffusivity in the crystallographic [100] and [010] directions parallel and perpendicular to a strain-induced foliation, respectively, as a function of temperature at atmospheric pressure on spinel lherzolites and spinel harzburgite. These rocks are considered representative for the subcontinental and suboceanic mantle, respectively. They found an anisotropy in thermal diffusivity on the order of 25 % with the maximum aligned in the direction of strain. Support of their experimental findings was provided by corroborating petrophysical modeling.

For mantle minerals, such as olivine and its high-pressure polymorphs, the β- and γ-spinels wadsleyite and ringwoodite, respectively, Xu et al. (2004) fitted phonon thermal diffusivity measured to 1 373 K and 20 GPa to a linear equation in pressure yielding a pressure coefficient of 0.036(4) GPa^{-1} (Table 5).

Osako et al. (2004) fitted thermal diffusivity measured to 1 100 K and 8.3 GPa on isotropic, single-crystal garnet and anisotropic olivine ($Fo_{93}Fa_7$; in three crystallographic directions) as upper and lower mantle constituents, respectively, to an exponential equation in pressure (Table 6). The pressure dependence of garnet and olivine was found to be on the order of 4 % GPa^{-1}–5 % GPa^{-1} and 3 % GPa^{-1}–4 % GPa^{-1}, respectively.

Thermal Storage and Transport Properties of Rocks, II: Thermal Conductivity and Diffusivity, Table 7 Coefficients for fitting the variation of thermal diffusivity (10^{-6} m^2 s^{-1}) with temperature (K) according to Equation 32 (Pertermann and Hofmeister, 2006)

Sample		Chemical composition	$a \times 10^6$ (m^2 s^{-1})	$b \times 10^6$ (m^2 s^{-1} K)	$c \times 10^6$ (m^2 s^{-1} K^2)	T_{max} (°C)
		Single crystals				
Olivines	Fo[001]	Mg_2SiO_4	0.3081	679.6	213 492	985
	FoCo[001]	$Mg_{1.99}Co_{0.01}SiO_4$	0.2347	587.8	172 482	1 477
	[010]	”	0.2415	115.4	165 515	1 181
	Needles[100]	$Mg_{1.84}Fe_{0.16}SiO_4$	0.7088	57.7	202 533	985
	[010]	”	0.3100	100.6	86 470	739
	[001]	”	0.3805	381.3	79 703	886
	Sumput[010]	$Mg_{1.87}Fe_{0.13}SiO_4$	0.3135	127.2	73 824	983
Sinhalite	[010]	$MgAlBO_4$	0.5546	−128.1	432 571	741
Chrysoberyl	[100]	$BeAl_2O_4$	0.5366	551.7	566 872	990
	[010]	”	0.3516	415.2	388 978	989
	[001]	”	0.6371	428.2	543 382	990
		Polycrystalline samples				
Dunites	#1	~$Mg_{1.8}Fe_{0.2}SiO_4$	0.2291	290.0	92 938	888
	#2	”	0.3563	178.4	93 356	1 083
Monticellite-bearing rock		$Ca_{1.15}$ $Mg_{0.79}Mn_{0.06}SiO_4$	0.3816	153.0	23 706	985
Hortonolite-bearing rock		$Mg_{1.2}Fe_{0.8}SiO_4$	0.2826	301.2	12 638	705
Fayalite-bearing slag		~$Fe_{1.98}Mn_{0.02}SiO_4$	0.2637	83.9	21 216	886
Fayalite-bearing rock		~$Fe_{1.84}Mn_{0.02}$ $Mg_{0.14}SiO_4$	0.1798	265.3	15 688	887

Fo forsterite, *FoCo* Co-doped forsterite

Variation with other factors

Micro-cracks and grain boundaries give rise to increased thermal resistance and to a reduction of the mean free path of radiation due to scattering of radiation. It is somewhat debated below which grain size the effect is negligible: Based on the agreement between diffusivities measured on minerals and rocks, Gibert et al. (2003) concluded that the effect of grain boundaries, thermal cracking, and secondary phases is negligible. Along the same lines, Seipold (1998) argued that grain size is much larger than the phonon mean free path and therefore grain boundaries should not interfere with heat diffusion. Branlund and Hofmeister (2008) confirm this for quartzites, but find diffusivities measured on chert, agate, and chalcedony to be lowered by grain boundaries. They propose that grain sizes above 1 μm should not affect heat transfer.

Additional advective heat transfer was identified by Seipold and Schilling (2003) due to the release of water adsorbed at the inner surfaces of voids (i.e., pores and cracks) in rocks at about 450 K and by dehydration of serpentinite at 850 K. Both processes create high local overpressures which are relieved by cracking if the overpressure exceeds the tensile strength of the rock. The resulting flow is then accompanied by a corresponding advective heat transfer. This phenomenon was observed and studied in laboratory experiments (Seipold and Schilling, 2003) but has implications for the lower crust and upper mantle with respect to recrystallization processes involving the discharge of fluids. These “crustal burps” provide the only conceivable way how some fluids from the mantle or lower crust my find their way to the Earth’s surface. The example discussed by Seipold and Schilling (2003) involves liberating water of crystallization during the conversion of serpentinite into forsterite and talc, followed by the formation of enstatite. As this involves heat advection as a separate heat transfer mechanism, this process is better addressed separately and not parameterized into some sort of “effective” heat transport property not directly linked to a physical process.

Summary

Understanding the thermal regime of the Earth requires appreciation of properties and mechanisms for storage, transport, and generation of heat with the Earth. Both experimental and indirect methods are available for inferring the corresponding rock properties. Steady-state heat conduction or transient heat diffusion is the dominant transport process in the Earth’s crust, except when appreciable fluid flow provides a mechanism for heat advection. For most crustal and mantle rocks, heat radiation sets in at temperatures above about 450 °C and becomes significant only at temperatures above 1 200 °C. At temperatures above 2 500 °C heat radiation becomes a dominant mechanism.

Acknowledgments

Extensive and constructive comments provided by an anonymous reviewer are greatly appreciated. Dr. Sukanta Roy kindly helped to meet the size limit of this contribution.

Bibliography

Aronsson, J. R., Bellottti, L. H., Eckroad, S. W., Emslie, A. G., McConnell, R. K., and von Thüna, P. C., 1970. Infrared spectra and radiative thermal conductivity of minerals at high temperatures. *Journal of Geophysical Research*, **75**(17), 3443–3456.

Beardsmore, G. R., and Cull, J. P., 2001. *Crustal Heat Flow*. Cambridge: Cambridge University Press.

Beck, A. E., 1988. Methods for determining thermal conductivity and thermal diffusivity. In Hanel, R., Rybach, L., and Stegena, L. (eds.), *Handbook of Terrestrial Heat Flow Density Determination*. Dordrecht: Kluwer, pp. 87–124.

Birch, F., 1942. Thermal conductivity and diffusivity. In Birch, F., Schairer, J. F., Spicer, H. C. (eds.), *Handbook of Physical Constants*, Special Paper 36. New York: Geological Society of America, pp. 243–266.

Blumm, J., and Lemarchand, S., 2002. Influence of test conditions on the accuracy of laser flash measurements. *High Temp.-High Pres.*, **34**, 523–528.

Branlund, J. M., and Hofmeister, A. M., 2008. Factors affecting heat transfer in natural SiO_2 solids. *Ameican Mineralogist*, **93**, 1620–1629.

Bruggeman, D. A. G., 1935. Berechnung verschiedener Konstanten von heterogenen Substanzen – I. Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper aus isotropen Substanzen. *Annalen der Physik*, **24**, 636–679.

Buntebarth, G., 1991. Thermal properties of the KTB-Oberpfalz VB core samples at elevated temperature and pressure. *Scientific Drilling*, **2**, 73–80.

Buntebarth, G., and Schopper, J. R., 1998. Experimental and theoretical investigations on the influence of fluids, solids and interactions between them on thermal properties of porous rocks. *Physics and Chemistry of the Earth*, **23**, 1141–1146.

Carmichael, R. S. (ed.), 1984. *CRC Handbook of Physical Properties of Rocks III*. Boca Raton: CRC Press.

Čermák, V., and Rybach, L., 1982. Thermal conductivity and specific heat of minerals and rocks. In Angenheister, G. (ed.), *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology,* New Series, V(1a). Berlin: Springer, pp. 305–343.

Clark, S. P., Jr., 1957. Radiative transfer in the Earth's mantle. *EOS. Transactions of the American Geophysical Union*, **38**, 931–938.

Clark, S. P., Jr., 1966. Thermal conductivity. In Clark, S. P., Jr. (ed.), *Handbook of Physical Constants,* Memoir 97. New York: Geological Society of America, pp. 459–482.

Clark, S. P., Jr., 1969. Heat conductivity in the mantle. In Hart, P. J. (ed.), *The Earth's Crust and Upper Mantle*, Geophysical Monograph 13. Washington, DC: American Geophysical Union, pp. 622–626.

Clauser, C., 1988. Opacity – the concept of radiative thermal conductivity. In Hänel, R., Rybach, L., and Stegena, L. (eds.), *Handbook of terrestrial heat flow density determination*. Dordrecht: Kluwer, pp. 143–165.

Clauser, C., 2006. Geothermal energy. In Heinloth, K. (ed.), *Landolt-Börnstein, Group VIII: Advanced Materials and Technologies*. Heidelberg-Berlin: Springer. Energy Technologies, Subvol. C: Renewable Energies, Vol. 3, pp. 480–595.

Clauser, C., 2009. Heat transport processes in the Earth's crust. *Surveys in Geophysics*, **30**, 163–191, doi:10.1007/s10712-009-9058-2.

Clauser, C., and Huenges, E., 1995. Thermal conductivity of rocks and minerals. In Ahrens, T. J. (ed.), *Rock Physics and Phase Relations – a Handbook of Physical Constants*. Washington, DC: American Geophysical Union. AGU Reference Shelf, Vol. 3, pp. 105–126.

Crain, E. R., 1986. *The Log Analysis Handbook*. Tulsa: Pennwell Publishing. Quantitative Log Analysis Methods Series, Vol. 1.

Davis, E. E., 1988. Oceanic heat-flow density. In Hänel, R., Rybach, L., and Stegena, L. (eds.), *Handbook of Terrestrial Heat Flow Density Determination*. Dordrecht: Kluwer, pp. 223–260.

Davis, G. M., Chapman, D. S., Van Wagoner, T. M., and Armstrong, P. A., 2007. Thermal conductivity anisotropy of metasedimentary and igneous rocks. *Journal of Geophysical Research*, **112**, B05216, doi:10.1029/2006JB004755.

Debye, P., 1914. Zustandsgleichung und Quantenhypothese mit einem Anhang über Wärmeleitung. In *Vorträge über die kinetische Theorie der Materie und der Elektrizität, Mathematische Vorlesungen an der Universität Göttingen,* VI. Leipzig: Teubner.

Diment, W. H., and Pratt, H. R., 1988. *Thermal Conductivity of Some Rock-Forming Minerals: A Tabulation*, Open File Report 88-690. Denver, CO: US Geological Survey.

Dreyer, W., 1974. *Materialverhalten Anisotroper Festkörper: Thermische und Elektrische Eigenschaften*. Wien: Springer.

Eucken, A., 1911. Über die temperaturabhängigkeit der wärmeleitfähigkeit fester nichtmetalle. *Annalen der Physik, 4. Folge*, **34**(2), 185–220.

Fricke, H., 1924. A mathematical treatment of the electric conductivity and capacity of disperse systems. *Physical Review*, **24**, 575–587.

Gibert, B., Seipold, U., Tommasi, A., and Mainprice, D., 2003. Thermal diffusivity of upper mantle rocks: influence of temperature, pressure, and the deformation fabric. *Journal of Geophysical Research*, **108**(B8), 2359, doi:10.1029/2002JB002108.

Goutorbe, B., Lucazeau, F., and Bonneville, A., 2006. Using neural networks to predict thermal conductivity from geophysical well logs. *Geophysical Journal International*, **166**(1), 115, doi:10.1111/j.1365-246X.2006.02924.x.

Hartmann, A., Rath, V., and Clauser, C., 2005. Thermal conductivity from core and well log data. *International Journal of Rock Mechanics and Mining Sciences*, **42**, 1042–1055, doi:10.1016/j.ijrmms.2005.05.015.

Hashin, Z., and Shtrikman, S., 1962. A variational approach to the theory of the effective magnetic permeability of multiphase materials. *Journal of Applied Physics*, **33**(10), 3125–3131.

Höfer, M., and Schilling, F. R., 2002. Heat transfer in quartz, orthoclase, and sanidine at elevated temperature. *Physics and Chemistry of Minerals*, **29**, 571–584.

Hofmeister, A. M., 1999. Mantle values of thermal conductivity and the geotherm from phonon lifetimes. *Science*, **283**, 1699–1706.

Hofmeister, A. M., 2005. Dependence of diffusive radiative transfer on grain-size, temperature, and Fe-content: Implications for mantle processes. *Journal of Geodynamics*, **40**, 51–72.

Hofmeister, A. M., 2006. Thermal diffusivity of garnets at high temperature. *Physics and Chemistry of Minerals*, **33**, 45–62.

Hofmeister, A. M., Branlund, J. M., and Pertermann, M., 2009. Properties of rocks and minerals – thermal conductivity of the Earth. In Price, G. D. (ed.), *Mineral Physics, Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 2, pp. 543–577.

Horai, K., 1971. Thermal conductivity of rock-forming minerals. *Journal of Geophysical Research*, **76**(5), 1278–1308.

Horai, K., 1991. Thermal conductivity of Hawaiian Basalt: a new interpretation of Robertson and Peck's data. *Journal of Geophysical Research*, **96**(B3), 4125–4132.

Horai, K., and Simmons, G., 1969. Thermal conductivity of rock-forming minerals. *Earth and Planetary Science Letters*, **6**(5), 359–368.

Kappelmeyer, O., and Haenel, R., 1974. *Geothermics – with special reference to application*. Berlin-Stuttgart: Bornträger.

Kobolev, V. P., Kutas, R. I., and Popov, Y. A., 1990. Method and results of research of thermal properties of Ural region rocks with

laser scanning. *Geophysical Journal*, **12**(4), 29–37, Naukova Dumka, Kiev (in Ukrainian).
Korvin, G., 1978. The hierarchy of velocity formulae: generalized mean value theorems. *Acta Geodaetica, Geophysica et Montanista*, **13**, 211–222.
Korvin, G., 1982. Axiomatic characterization of the general mixture rule. *Geoexploration*, **19**, 267–276.
Mottaghy, D. C., Schellschmidt, R., Popov, Y. A., Clauser, C., Kukkonen, I. T., Nover, G., Milanovsky, S., and Romushkevich, R. A., 2005. New heat flow data from the immediate vicinity of the Kola super-deep borehole: vertical variation in heat flow confirmed and attributed to advection. *Tectonophysics*, **401**(1–2), 119–142, doi:10.1016/j.tecto.2005.03.005.
Osako, M., Ito, E., and Yoneda, A., 2004. Simultaneous measurement of thermal conductivity and thermal diffusivity for garnet and olivine under high pressure. *Physics of the Earth and Planetary Interiors*, **143–144**, 311–320.
Parker, W. J., Jenkins, R. J., Butler, C. P., and Abbott, G. I., 1961. Flash method of determining thermal diffusivity, heat capacity, and thermal conductivity. *Journal of Applied Physics*, **32**(9), 1679–1684.
Peierls, R., 1929. Zur kinetischen Theorie der Wärmeleitung in Kristallen. *Annalen der Physik, 5. Folge*, **3**, 1055–1101.
Pertermann, M., and Hofmeister, A. M., 2006. Thermal diffusivity of olivine-group minerals at high temperature, *American Mineralogist*, **91**, 1747–1760.
Petrunin, G. I., Popov, V. G., and Il'in, I. A., 2004. Conductive heat transfer in Plagioclases. *Izvestiya, Physics of the Solid Earth (English Translatrion)*, **40**(9), 752–759.
Popov, Y. A., and Mandel, A. M., 1998. Geothermal study of anisotropic rock masses. *Izvestiya, Physics of the Solid Earth (English Translation)*, **34**(11), 903–915.
Popov, Y. A., Berezin, V. V., Solov'yev, G. A., Romushkevich, R. A., Korostelev, V. M., Kostyurin, A. A., and Kulikov, I. V., 1987. Thermal conductivity of minerals. *Izvestiya, Physics of the Solid Earth (English Translation)*, **23**(3), 245–253.
Popov, Y. A., Pevzner, L. A., Romushkevich, R. A., Korostelev, V. M., and Vorob'ev, M. G., 1995. Thermophysical and geothermal sections obtained from Kolvinskaya well logging data. *Izvestiya, Physics of the Solid Earth (English Translation)*, **30**(9), 778–789.
Popov, Y. A., Pevzner, L. A., Pimenov, V. P., and Romushkevich, R. A., 1999a. New geothermal data from the Kola superdeep well SG-3. *Tectonophysics*, **306**(3–4), 345–366.
Popov, Y. A., Pribnow, D. F. C., Sass, J. H., Williams, C. F., and Burkhardt, H., 1999b. Characterization of rock thermal conductivity by high-resolution optical scanning. *Geothermics*, **28**, 253–276.
Popov, Y., Tertychnyi, V., Romushkevich, R., Korobkov, D., and Pohl, J., 2003. Interrelations between thermal conductivity and other physical properties of rocks: experimental data. *Pure and Applied Geophysics*, **160**, 1137–1161.
Ray, L., Förster, H.-J., Schilling, F. R., and Förster, A., 2006. Thermal diffusivity of felsic to mafic granulites at elevated temperatures. *Earth and Planetary Science Letters*, **251**, 241–253.
Reibelt, M., 1991. *Study on the Influence of Surface Structure and Fluid Saturation of Rocks on the Determination of Thermal Conductivity by a Half-space Line Source*. Unpublished diploma thesis (in German). Germany: Institut für Angewandte Geophysik, Petrologie und Lagerstättenforschung, Technische Universität Berlin.
Rosseland, S., 1930. The principles of quantum theory. In Eberhard, G., Kohlschütter, A., and Ludendorff, H. (eds.), *Handbuch der Astrophysik*. Berlin: Springer, Vol. 3(1), pp. 452–474.
Roy, R. F., Beck, A. E., and Touloukian, Y. S., 1981. *Thermophysical Properties of Rocks*. In Touloukian, Y. S., Judd, W. R., and Roy, R. F. (eds.), *Physical Properties of Rocks and Minerals*, McGraw-Hill/CINDAS Data Series on Material Properties, Vol. II-2. New York: McGraw-Hill, pp. 409–502.
Sass, J. H., Lachenbruch, A. H., Moses, T. H., Jr., and Morgan, T., 1992. Heat flow from a scientific research well at Cajon Pass, California. *Journal of Geophysical Research*, **97**(B4), 5017–5030.
Schärmeli, G., 1982. Anisotropy of olivine thermal conductivity at 2.5 GPa up to 1500 K measured on optically non-thick sample. In Schreyer, W. (ed.), *High-Pressure Researches in Geoscience*. Schweizerbart, Stuttgart: Germany, pp. 349–373.
Schilling, F. R., 1999. A transient technique to measure thermal diffusivity at elevated temperatures. *European Journal of Mineralogy*, **11**, 1115–1124.
Schulz, B., 1981. Thermal conductivity of porous and highly porous materials. *High Temperatures-High Pressures*, **13**, 649–660.
Seipold, U., 1998. Temperature dependence of thermal transport properties of crystalline rocks – a general law. *Tectonophysics*, **291**(1–4), 161–171.
Seipold, U., and Schilling, R. R., 2003. Heat transport in serpentinites. *Tectonophysics*, **370**, 147–162.
Shankland, T. J., Nitsan, U., and Duba, A. G., 1979. Optical absorption and radiative heat transport in olivine at high temperature. *Journal of Geophysical Research*, **84**(B4), 1603–1610.
Somerton, W. H., 1992. *Thermal Properties and Temperature Related Behavior of Rock/Fluid Systems*. Amsterdam: Elsevier.
Tommasi, A., Gibert, B., Seipold, U., and Mainprice, D., 2001. Anisotropy of thermal diffusivity in the upper mantle. *Nature*, **411**, 783–786.
Vosteen, H.-D., and Schellschmidt, R., 2003. Influence of temperature on thermal conductivity, thermal capacity, and thermal diffusivity for different types of rock. *Physics and Chemistry of the Earth*, **28**(9–11), 499–509.
Whittington, A. G., Hofmeister, A. M., and Nabelek, P. I., 2009. Temperature-dependent thermal diffusivity of the Earth's crust and implications for magmatism. *Nature*, **458**, 319–321, doi:10.1038/nature07818.
Williams, C. F., and Anderson, R. A., 1990. Thermophysical properties of the Earth's Crust: In situ measurements from continental and ocean drilling. *Journal of Geophysical Research*, **95**(B6), 9209–9236.
Xu, Y., Shankland, T. J., Linhardt, S., Rubie, D. C., Langenhorst, F., and Klasinski, K., 2004. Thermal diffusivity and conductivity of olivine, wadsleyite and ringwoodite to 20 GPa and 1373 K. *Physics of the Earth and Planetary Interiors*, 143–144, 321–336.
Zimmerman, R. W., 1984. *The Effect of Pore Structure on the Pore and Bulk Compressibilities of Consolidated Sandstones*. Ph.D. thesis. Berkeley, CA: University of California.
Zimmerman, R. W., 1989. Thermal conductivity of fluid-saturated rocks. *Journal of Petroleum Science and Engineering*, **3**(3), 219–227.
Zoth, G., and Hänel, R., 1988. Appendix. In Hänel, R., Rybach, L., and Stegena, L. (eds.), *Handbook of Terrestrial Heat Flow Density Determination*. Dordrecht: Kluwer, pp. 449–466.

Cross-references

Heat Flow, Continental
Heat Flow, Seafloor: Methods and Observations
Thermal Storage and Transport Properties of Rocks, I: Heat Capacity and Latent Heat

TIME REVERSAL IN SEISMOLOGY

Carène Larmat[1], Clarence S. Clay[2]
[1]Geophysics Group, EES-17, MS D443, Los Alamos National Laboratory, Los Alamos, NM, USA
[2]Marine Geophysics and Wave Propagation, Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA

Definition

Time-Reversal – This refers to a collection of techniques that aim to focus wave energy onto a specific point in space and time, implicitly using the *source-receiver reciprocity* and the *invariance* of the wave propagation equation to the transformation $t \rightarrow -t$. The usual way to perform a time-reversal experiment is to simultaneously rebroadcast from a set of receivers the recorded signals read backward. By doing this, an approximate and reverse version of an original wave propagation movie is created, which will display wavefronts converging back onto the source point. A variant of this procedure is *reciprocal Time-Reversal*, where the read backward signals are rebroadcast from the source point instead of from the recording point. Based on the *source-receiver reciprocity*, the wavefield will in this case focus on the receiver.

Invariance of the wave equation – In the absence of loss terms, the wave equation only contains second-order space and time derivatives. This implies that if u(x,t) is a wavefield solution, then u(x,−t), u(−x,t), and u(−x,−t) are also solutions of the equation. This means that the propagation direction of a wavefield may be reversed if it has been measured at any point of space at a given moment. Consideration about the correct boundary conditions to be applied can be found in the subsequent references.

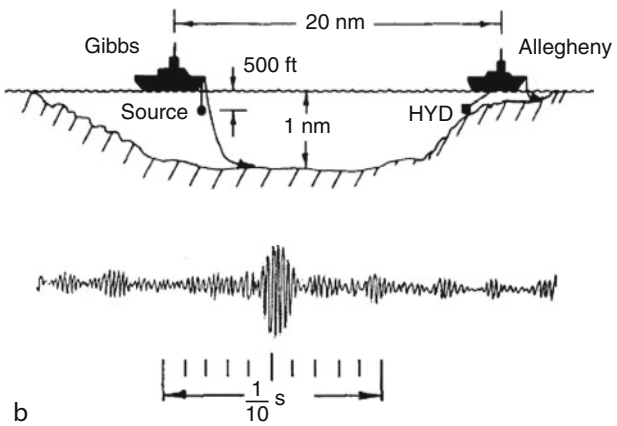
Source-receiver reciprocity – The principle of reciprocity states that a source and receiver may be interchanged and the same waveform will be observed. The conditions required are that the considered signal is transient, nonlinear effects are negligible, and that the medium remains unchanged.

History

Imagine filming the ripples formed when a pebble is dropped into a pond, and playing the movie backward. You will observe the ripples converge back onto the original impact point, indicating its location by a local concentration of wave energy. The history of developing Time-Reversal is strongly related to the ability to physically or numerically realize this thought experiment based on progress in understanding wave propagation and in developing the necessary tools.

The story of Time-Reversal starts in the 1950–1960s in a cluttered room of the Hudson Laboratory of Columbia University where Antares Parvulescu (1923–1998) tested his idea of time-reversed transmissions. A microphone was placed about 25 ft from a loud speaker, from which a series of short impulses were sent 1 s apart. A sequence of the signal due to the room reverberations was received at the microphone and tape-recorded. The tape was removed and played backward. These "time-reversed signals" (called then "matched signals") were played by the loud speaker (notice it is a reciprocal Time-Reversal experiment that was carried out). When an indiscernible rushing noise could be heard over the rest of the room, the scope at the microphone displayed a signal that looked like the autocorrelation of the reverberation signal. When we indeed brought our ears to the microphone, the rushing noise became a simple "click," indicating that localization in space and time of the wave energy did occur! This result was first published in a terse abstract (Parvulescu, 1961) and is further documented in Parvulescu (1995) and in chapter 7 of the book of Tolstoy and Clay (1966 and its revision in 1987).

In 1961–1962, Parvulescu and colleagues decided to take this matched signal experiment to sea. This was a serious logistic and scientific challenge as most of the geophysicists at that time did not believe that transmission in ocean was reproducible. The marine experiment used two oceanographic research ships, the Gibbs and Allegheny, with deep anchoring capability in the Tongue of Ocean, which is an oceanic trench. Figure 1a shows the setting of the experiment. The received signals at the Allegheny, lengthened by multiple reverberations, were radioed back to the Gibbs to be then transmitted from the source (this was again a reciprocal Time-Reversal experiment). Figure 1b shows the peak of pressure created on the hydrophone. The report of this first "field" Time-Reversal experiment is documented in Parvulescu and Clay (1965). Details of further experimentations can be found in chapters 4–8 of Tolstoy and Clay (1966–1987).



Time Reversal in Seismology, Figure 1 **(a)** Ship positions and instruments used for the marine experiment. Gibbs has the source, and Allegheny has the receiver (HYD). **(b)** Signal received at the hydrophone when performing a reciprocal time-reversal between the two ships in an oceanic trench. Figures from Parvulescu and Clay (1965) and Tolstoy and Clay (1966).

In 1980s, Time-Reversal was first used in seismology by George McMechan and colleagues at the University of Texas. They modified techniques developed for exploration geophysics to image the primary source of excitation instead of geologic reflectors and rediscovered the principle of Time-Reversal. They successfully imaged the source of a small 1983 aftershock earthquake in California using a purely acoustical 2D model to numerically rebroadcast signals recorded at a very dense network in a 15 by 12 km vertical slice of the formation (McMechan, 1982).

In the late 1980s, advances in microelectronics, storage capacity, and automated signal processing led to the development of the *Time-Reversal Mirror* (TRM) by Mathias Fink and colleagues at the Laboratoire Ondes et Acoustique (LOA) of the University of Paris VII (Fink, 1992). TRMs are finite-size arrays of transmit-receive transducers that are at the base of active research on theoretical limit of Time-Reversal (Fink et al., 2000; Fink, 2006). TRMs have made possible novel applications of Time-Reversal ranging from nondestructive evaluation (Ulrich et al., 2008a, b) to medical imaging (Fink and Tanter, 2010).

For references and more technical details about Time-Reversal, see the reviews by Fink et al., 2000; Fink, 2006; Anderson et al., 2008; and Larmat et al., 2010.

Time-reversal and complexity

The Earth and oceans are complex environments in which transmitted signals quickly become difficult to interpret, for example, instead of a single pulse for the first arrival, the P-wave is accompanied by a coda train due to the scattering within the crust. Contrary to many location methods based on simple paradigms, Time-Reversal thrives with complexity. In a homogeneous and purely acoustic medium, only one phase will be recorded on a single receiver from a source point. Introduce a reflecting interface and two phases will be recorded: the primary and a copy of the source impulse due to the reflection on the interface. The two combined phases contain more information about the source location than the primary phase alone. All these different phases will converge and add up onto the source point when traveling backward. Several studies (e.g., Dowling and Jackson, 1992; Fink et al., 2000; Blomgren et al., 2002) have actually demonstrated the width of the focus is smaller in randomly heterogeneous media than in the homogeneous case (implying better resolution on the location). This is referred as *super-resolution.* Theoretical discussion about Time-Reversal resolution can be found in Fink et al. (2000). A second important point for complex media is that any arbitrary time-segment will return to the source. Time-Reversal does not need any interpretation of the signals.

These unique facets of Time-Reversal explain the wide use of the method in very different disciplines handling with complex media, for example, target detection method and communication in underwater acoustics, nondestructive evaluation (i.e., location of defects in intact specimens), improved imaging and destruction of stones/tumors in medicine, and source location and imaging techniques in geophysics.

Location of seismic sources

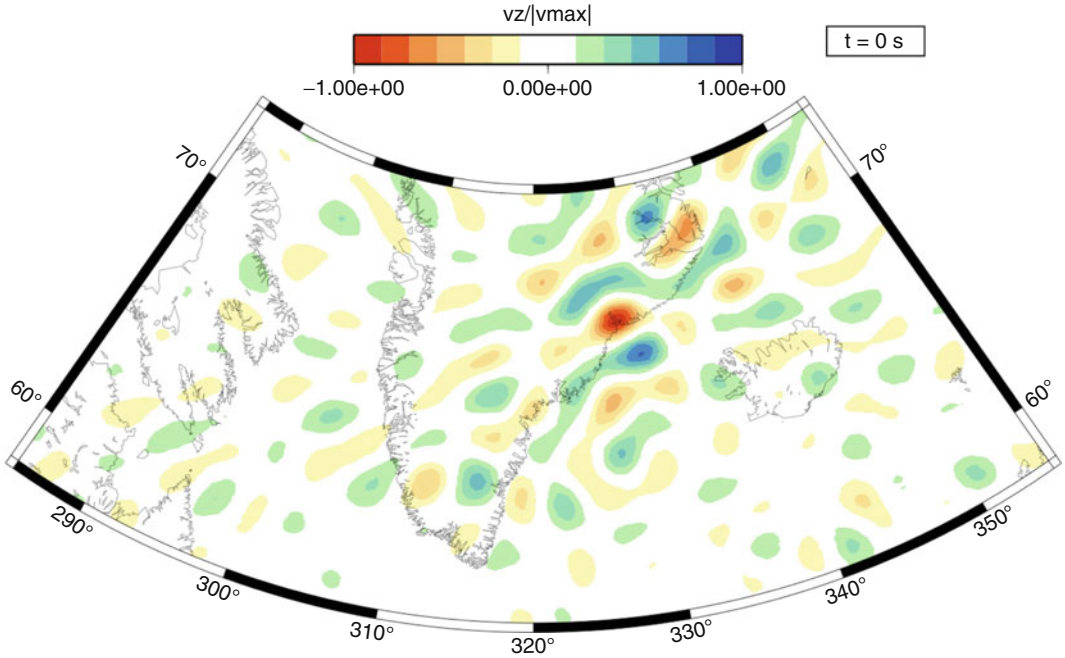
Since the late 1990s, several groups have performed Time-Reversal location of seismic events (for references, see Anderson et al., 2008; Larmat et al., 2010), harnessing the developing computing power. These applications have demonstrated that contrary to what was thought by McMechan, Time-Reversal does work with "sparse" rebroadcasting networks (i.e., with receivers way more than one wavelength apart). Figure 2 shows the time-reversal result from a 2001 magnitude-5 glacial earthquake in Greenland by Larmat et al. (2008). Glacial earthquakes were discovered in 2003 by Ekström et al. (2003) as long-period signal barely above the noise level. The signal from 146 stations in the Northern hemisphere was filtered between 0.01 and 0.02 Hz, time-reversed and rebroadcast into a 3D Earth model. The figure shows the snapshot of the vertical component of the time-reversed displacement at the moment of the focus (actually the vertical component normalized by the maximum value reached on the snapshot). The focus appears as a two-lobe pattern of opposite polarity indicating the origin of the seismic signal. The two-lobe pattern reveals that the nature of the source excitation was a single force vector, which is consistent with recent source analysis by Tsai et al. (2008). Had the source been a simple monopole, the focus would have manifested itself as a single point (see details in Larmat et al., 2008).

In the last 5 years, Time-Reversal application to seismology has evolved into using 3D full-waveform modeling to locate any type of seismic sources (e.g., point source, long lasting, finite) with "no specific interpretation . . . made of the various arrivals" as pointed out by Brian Kennett (1983). One application can be location of the origin of various emergent signals such as tremors (Steiner et al., 2008; Larmat et al., 2009; Lokmer et al., 2009) and weaker microseismicity. It must be noted that several Time-Reversal source location have been carried out of late.

Other location methods are based on the core idea of sending back some recorded signals to a given point in space and time. All back-projection related methods can be seen as simplified Time-Reversal procedures for which the backward propagation is reduced to shifting the time series by predicted arrival times (e.g., Source Scanning Algorithm of Kao et al. (2005).

Location of features in the velocity model

Time-reversed propagation is also used to image the subsurface structure. The exploration methods used by McMechan have evolved into reverse-time migration (e.g., Baysal et al.,1983). Next generation tomography



Time Reversal in Seismology, Figure 2 One hundred and forty-six vertical displacement signals of a glacial earthquake were time-reversed and sent back in a 3D Earth model. The vertical component (vz) of the created time-reversal signal at the moment of the focus displays a two-lobe pattern of opposite polarity consistent with a single force vector as the source mechanism (adapted from Larmat et al., 2008).

models are currently created based on the adjoint method that uses a forward and a backward propagation to locate the source of discrepancies between data and predicted waveforms (Tromp et al., 2005).

Time-reversal and correlation methods

The signal produced at point A from source S can be expressed with the Green function formalism: $S_A(t) = G(x_S, x_A, t) \otimes s(t)$ where *s(t)* is the source time function and $\otimes$ denotes the convolution operation. If you time-reverse the signal $S_A(t)$, and rebroadcast it from A, the signal produced on point S will be:

$$S_S^{TR}(t) = G(x_A, x_S, t) \otimes S_A(T-t) = [G(x_S, x_A, T-t) \otimes G(x_S, x_A, -(T-t)] \otimes s(t) \quad (1)$$

where the reciprocity between A and S of the Green function has been used. The right term of this expression is similar to the expression of the autocorrelation of S_A in terms of Green functions. More generally, the cross-correlation of the signal recorded at a point A and B is:

$$C_{AB}(t) = [G(x_S, x_A, t) \otimes G(x_S, x_B, -t)] \otimes [s(T-t) \otimes s(t)] \rightarrow S[G(x_A, x_B, t) + G(x_A, x_B, -t)] \quad (2)$$

which states that under certain conditions the cross-correlation of two signals is composed of the superposition of the Green function between the two receivers and its time-reversed version. Equation (2) is the basis of what is known as *interferometry* or imaging with ambient noise, as the Green function contains information about the properties of the medium (see *Seismic, Ambient Noise Correlation*, the review by Snieder and Wapenaar, 2010 and the pioneer paper of Claerbout, 1968).

To come back to Time-Reversal, if the source time function is considered to be a dirac, the time-reversed signal at the original point source S is:

$$S_S^{TR}(t) = [G(x_S, x_A, T-t) \otimes G(x_S, x_A, -(T-t)] \quad (3)$$

This expression is typical of a *matched filter*, whose output is optimal in some sense. Whatever the Green function, the Time-Reversal wavefield is maximal at the focus time and at the source point (Fink et al., 2000, p. 1946). Applications derived from this interpretation of Time-Reversal consist in locating seismic sources, such as trapped miners, by comparing their transmitted distress signals with a database of prerecorded Green functions (created by recording the seismic response for each possible source location) (Hanafi et al., 2009).

Summary

Interest in seismic source study has recently shifted from event-based temporary observation to continuous

monitoring using dense seismic networks. Seismologists now study a wealth of signals due to less energetic sources than earthquakes such as tremors and microseismicity. Glacial earthquakes, landslides, Earth Hum, and meteorological phenomena such as hurricanes belong to a group of events that defines "environmental seismology," a term coined by the French seismologist Jean-Paul Montagner.

Microseismic noise is now used for imaging the Earth. Due to their emergent nature, these signals represent a challenge to classical location methods based on triangulation. Time-Reversal as a method and Time-Reversal ideas are emerging as a promising framework to foster future location and imaging methods. The success of Time-Reversal as a source location method shall rely on the coverage of seismic networks, the accuracy of high-resolution velocity models, and the performance of numerical schemes to be used for the backward propagation. Still unclear are the convergence and resolution power of Time-Reversal under various conditions. Our ability to reliably observe and model scattering phenomena is expected to be an important part of the answer.

Acknowledgment

We gratefully acknowledge the support of the U.S. Department of Energy through the LANL/LDRD Program for this work.Carène Larmat & Clarence S. Clay.

Bibliography

Anderson, B., Griffa, M., Larmat, C., Ulrich, T. J., and Johnson, P. A., 2008. Time reversal. *Acoustics Today*, **4**(1), 5.

Baysal, E., Kosloff, D. D., and Sherwood, J. W. C., 1983. Reverse time migration. *Geophysics*, **48**(11), 1514–1524.

Blomgren, P., Papanicolaou, G., and Zhao, H., 2002. Super-resolution in time-reversal acoustics. *The Journal of the Acoustical Society of America*, **111**(1), 230–248.

Claerbout, J. F., 1968. Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, **33**, 264–269.

Dowling, D. R., and Jackson, D. R., 1992. Narrow-band performance of phase-conjugate arrays in dynamic random media. *The Journal of the Acoustical Society of America*, **91**, 3257–3277.

Ekström, G., Nettles, M., and Abers, G., 2003. Glacial earthquakes. *Science*, **302**, 622–624.

Fink, M., 1992. Time-reversal of ultrasonic fields–Part I: basic principles. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **39**(5), 555–566.

Fink, M., 2006. Time-reversal acoustics in complex environments. *Geophysics*, **71**(4), SI151–SI164.

Fink, M., and Tanter, M., 2010. Multiwave imaging and super resolution. *Physics Today*, **63**(2), 28–33.

Fink, M., Cassereau, D., Derode, A., Prada, C., Roux, P., Tanter, M., Thomas, J.-L., and Wu, F., 2000. Time-reversed acoustics. *Reports on Progress in Physics*, **63**(12), 1933–1995.

Hanafi, S. M., Cao, W., McCarter, K., and Schuster, G. T., 2009. Using super-stacking and super-resolution properties of time-reversal mirrors to locate trapped miners. *The Leading Edge*, **28**(3), 302–307.

Kao, H., Shan, S.-J., Dragert, H., Rogers, G., Cassidy, J. F., and Ramachandran, K., 2005. A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature*, **436**, 841–844.

Kennett, B. L. N., 1983. A new way to estimate source parameters. *Nature*, **302**, 659–660.

Larmat, C., Tromp, J., Liu, Q., and Montagner, J.-P., 2008. Time reversal location of glacial earthquakes. *Journal of Geophysical Research*, **113**, B09314.

Larmat, C. S., Guyer, R. A., and Johnson, P. A., 2009. Tremor source location using time reversal: selecting the appropriate imaging field. *Geophysical Research Letters*, **36**, L22034.

Larmat, C. S., Guyer, R. A., and Johnson, P. A., 2010. Time-reversal methods in geophysics. *Physics Today*, **63**, 31–35.

Lokmer, B., O'Brien, G. S., Stich, D., and Bean, C. J., 2009. Time reversal imaging of synthetic volcanic tremor sources. *Geophysical Research Letters*, **36**, L12308.

McMechan, G. A., 1982. Determination of source parameters by wavefield extrapolation. *Geophysical Journal of the Royal Astronomical Society*, **71**, 613–628.

Parvulescu, A., 1961. MESS processing. *The Journal of the Acoustical Society of America*, **33**, 1674.

Parvulescu, A., 1995. Matched-signal (MESS) processing by the ocean. *The Journal of the Acoustical Society of America*, **98**, 943–960.

Parvulescu, A., and Clay, C. S., 1965. Reproducibility of signal transmissions in the ocean. *Radio Engineering and Electronic Physics*, **29**(4), 223–228.

Snieder, R., and Wapenaar, K., 2010. Imaging with ambient noise. *Physics Today*, **63**(9), 44–49.

Steiner, B., Saenger, E. H., and Schmalholz, S. M., 2008. Time reverse modeling of low-frequency microtremors: application to hydrocarbon reservoir localization. *Geophysical Research Letters*, **35**, L03307.

Tolstoy, I., and Clay, C. S., 1966. *Ocean Acoustics, Theory and Experiment in Underwater Sound*. McGraw-Hill (Translated into Russian and published, 1969. Revised-edition 1987, the Acoustical Society of America – the American Institute of Physics).

Tromp, J., Tape, C., and Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels. *Geophysical Journal International*, **160**, 195–216.

Tsai, V. C., Rice, J. R., and Fahnestock, M., 2008. Possible mechanisms for glacial earthquakes. *Journal of Geophysical Research*, **113**, F03014.

Ulrich, T. J., Sutin, A. M., Guyer, R. A., and Johnson, P. A., 2008a. Time reversal and non-linear elastic wave spectroscopy (TR NEWS) techniques. *International Journal of Non-Linear Mechanics*, **43**, 209–216.

Ulrich, T. J., Sutin, A. M., Claytor, T., Pallin, P., Le Bas, P.-Y., and TenCate, J. A., 2008b. The time reversal elastic nonlinearity diagnostic applied to the evaluation of diffusion bonds. *Applied Physics Letters*, **93**, 151914.

Cross-references

Earthquake, Focal Mechanism
Earthquake, Location Techniques
Propagation of Elastic Waves: Fundamentals
Seismic Data Acquisition and Processing
Seismic Tomography
Seismic, Ambient Noise Correlation
Seismic, Migration
Seismic, Waveform Modeling and Tomography
Seismological Networks

TRAVELTIME TOMOGRAPHY USING CONTROLLED-SOURCE SEISMIC DATA

Colin A. Zelt
Department of Earth Science, Rice University, Houston, TX, USA

Synonyms

Active-source data; Traveltime inversion

Definition

Traveltime. The total time for a particular wave to travel from a source to a receiver.
Tomography. Determination of model structure by back projection of the data along a path connecting a source and a receiver.
Controlled-source seismology. Determination of earth structure using seismic waves generated by artificial sources such as chemical explosives, mechanical vibrators, weight drops, gun shots, etc.

Background

Traveltime tomography is the main method by which the Earth's seismic velocity structure is determined on all scales, from the upper few meters to the whole mantle. It was adapted from algorithms used in medical imaging in the 1970s (Dziewonski and Anderson, 1984). In seismology, the term tomography refers to the back projection of data along a path connecting a source and receiver using a mathematical inverse method. Ideally, the traveltimes of seismic arrivals corresponding to many criss-crossing rays are used to construct a two-dimensional (2D) or 3D image of the Earth's seismic velocity variations. Tomography is a type of inverse problem, although the terms traveltime inversion and traveltime tomography are used in the seismological community without a well-defined distinction. The latter usually implies a uniform, fine grid model parameterization, in which a smooth model is sought. Any departure from this approach is usually referred to as traveltime inversion, although the distinction is somewhat arbitrary (Levander et al., 2007); this article will discuss both approaches. This article concerns traveltime tomography using controlled-source data and does not discuss other types of tomography that use seismic amplitudes, phases, or waveforms as input data, nor earthquake or surface-wave data.

Data from earthquakes and artificial (controlled) sources are used in traveltime tomography, the former usually for crustal or mantle structure, the latter usually for crustal or near-surface structure. Much of the theory and practice of traveltime tomography is the same or similar for controlled-source and earthquake data. However, earthquake tomography usually uses delay times and includes a determination of the source locations (hypocenters), whereas controlled-source tomography usually uses total times and the source locations are known very accurately. Also, the degree of lateral velocity heterogeneity in the crust, and in particular the near-surface, can be much greater than in the mantle, so there is often no sense of a reference model in controlled-source seismology as there is in earthquake seismology, and this in turn means that a nonlinear tomographic method must be applied to controlled-source data in which the rays must be determined along with the unknown velocity structure as part of the inverse problem. Finally, controlled-source data are often acquired using sources and receivers distributed along a straight or nearly straight line, yielding a so-called 2D dataset. In this case, 2D modeling is used in which lateral homogeneity perpendicular to the line is assumed. Earthquake tomography is almost always 3D given the typical distribution of earthquakes and stations. Today, 3D controlled-source data is not uncommon, but it is expensive to acquire, especially for crustal-scale studies. In some studies, both earthquake and controlled-source data have been inverted simultaneously (e.g., Ramachandran et al., 2005).

Controlled-source seismology can be divided into two main approaches: (1) near-vertical reflection methods, and (2) refraction/wide-angle reflection methods. Traveltime tomography is applied to data from both experiments, although it is more common for refraction data. The most common applications of traveltime tomography using controlled-source data are (1) tectonic studies of the crust and uppermost mantle, (2) earthquake risk studies for upper crustal fault and basin geometries, (3) petroleum exploration in crosswell studies, model building to facilitate migration, and for refraction statics, and (4) environmental/engineering studies of the near surface (as shallow as a few meters). The traveltimes utilized are often limited to those of the first arrivals corresponding to direct, refracted (turning) or diffracted waves. Sometimes the traveltimes of primary reflections are included, for example, from the basement overlying sediments or the crust-mantle (Moho) boundary, necessitating the need for layers or interfaces in the model parameterization. Only rarely are later refracted arrivals, multiples or conversions utilized. Traveltimes are picked from the seismograms usually either interactively by eye, or using a semi-automated scheme whereby a few picks are made interactively, and the intervening picks are determined automatically using a cross-correlation scheme (Zelt, 1999). Usually only very high quality, spatially-dense data can be picked using fully automated schemes.

Most controlled-source tomography uses only P-wave traveltimes to constrain the compressional wave-velocity structure because P wave are easiest to generate and therefore the P wave arrivals are usually the strongest and easiest to pick. However, S wave experiments are quite common in environmental/engineering studies (e.g., Chambers et al., 2009). Theoretically, traveltime tomography is the same for P- and S-wave data, and if both data are available in the same study, Poisson's ratio can be estimated. The forward modeling component of traveltime tomography is almost always based on ray theory, an infinite-frequency approximation of wave propagation, although recently a finite-frequency approach for

controlled-source data has been proposed (Zelt, 2009). Almost all applications of traveltime tomography assume isotropic media because it is rarely necessary to invoke anisotropy to model the observed data. Exceptions to this are mainly for crosswell data and in studies of the oceanic crust (e.g., Pratt and Chapman, 1992; Dunn et al., 2000). Traveltime tomography is applied to land and marine data, as well as onshore-offshore data, using the same tomographic methodology. Arrays of airguns are the most common marine source, while on land chemical explosives, mechanical vibrators, weight drops, hammers, shotguns and rifles are common sources.

The ideal dataset for traveltime tomography is one with as dense a spacing of sources and receivers as possible, along a line for a 2D experiment, or areally distributed for a 3D experiment. As a result, a subsurface point will be sampled by rays at as many different angles as possible, which in turn will yield the best possible spatial model resolution as discussed later. There is usually very little processing of seismic data before traveltime picking, only that required to "clean up" the data to make it easier for picking. This may include bandpass filtering, trace editing, velocity filtering, trace mixing or binning, or deconvolution. Picks should not be interpolated to provide uniform spatial coverage if there are significant trace gaps, since this will provide an incorrect sense of model constraint (Zelt, 1999). Uncertainties should be assigned to the pick times to avoid over- or under-fitting the data, and to allow the appropriate up- and down-weighting of less noisy and more noisy data, respectively (Zelt, 1999). Elevation or bathymetric corrections should be avoided as they depend on the unknown velocity structure. Instead, the known surface topography or bathymetry should be incorporated into the 2D or 3D model. For a 2D experiment in which the shot and receiver locations deviate significantly from a straight line, it may be necessary to perform 2.5D modeling, i.e., using a 3D model and 3D ray tracing, and the known topographic or bathymetric surfaces, but keeping the velocity model homogeneous in a direction perpendicular to the main trend of the profile (e.g., Van Avendonk et al., 2004). Given the spatial resolution of most velocity models derived from traveltime tomography, the degree of profile "crookedness" must be substantial to warrant 2.5D modeling (Schmelzbach et al., 2008; Figure 1).

There are several review papers relevant to this article. Nowack and Braile (1993) review traveltime tomography methods in 1D and 2D media. Zelt (1999) focuses on 2D traveltime tomography and inversion, as well as model assessment methods. Rawlinson and Sambridge (2003a) present an overview of traveltime tomography for 2D and 3D structure. Levander et al. (2007) review all aspects of crustal-scale controlled-source seismology, including traveltime tomography.

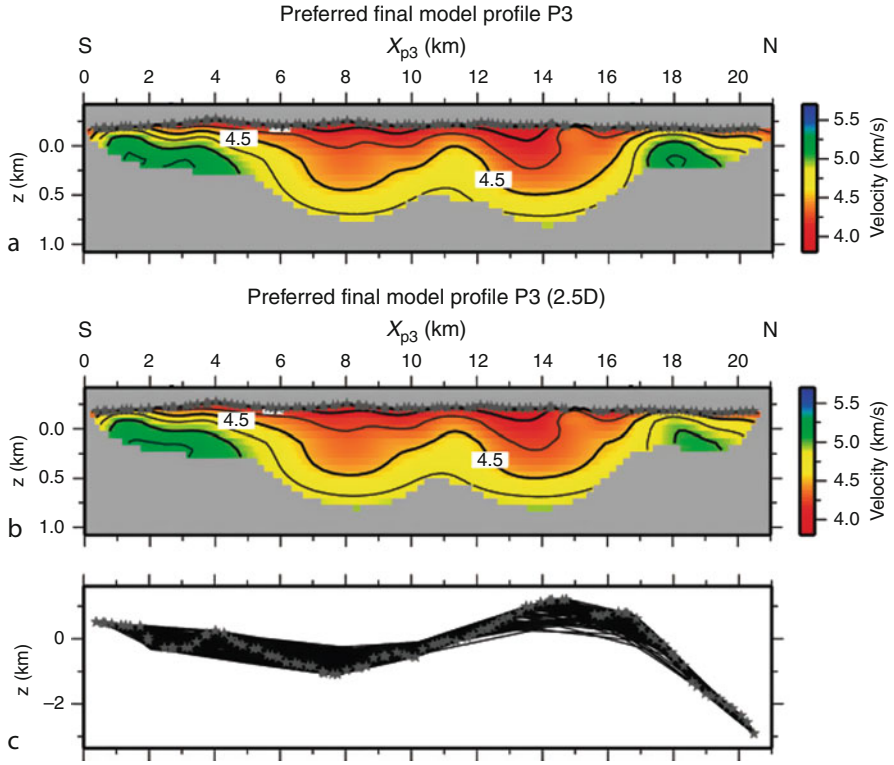
Forward modeling

In the 1970s, ray theory, an infinite-frequency approximation of wave propagation, for laterally heterogeneous media was developed (Cerveny et al., 1977, McMechan and Mooney, 1980). This allows earth models of arbitrary 2D and 3D complexity to be considered, provided the tenants of ray theory are honored, that is, the velocity field varies slowly with respect to the seismic wavelengths. Sharp velocity changes are modeled as velocity discontinuities (layer boundaries) using Snell's law. As a result, ray theory models are typically composed of one or a few layers, bounded by smoothly varying interfaces, within which the velocity varies smoothly. Ray theory forms the basis for the forward modeling used today in almost all traveltime tomography algorithms. There are two main types of infinite-frequency modeling: ray tracing and wavefront tracking.

Two end-member approaches to ray tracing are used: numerical and analytical. In the former case, the velocity field is specified by a regular (e.g., McMechan and Mooney, 1980) or irregular (e.g., Zelt and Ellis, 1988) grid of nodes, together with a scheme for interpolation between nodes, and a ray is traced by specifying a take-off angle from a point source and solving a set of ordinary differential equations, the ray tracing equations (Cerveny et al., 1977). In the analytic case, the velocity field is specified by a regular (e.g., Chapman and Drummond, 1982; White, 1989; Rawlinson et al., 2001) or irregular (e.g., Spence et al., 1984) grid of nodes, but with a form of interpolation, for example linear, such that the rays can be calculated analytically within each cell of the grid. In practice, the numerical and analytical techniques may be about equally efficient if the analytical method uses a fine grid of nodes to represent a sufficiently smooth velocity field as required by ray theory, and if the numerical scheme adjusts the step length according to the velocity gradient (Zelt and Ellis, 1988).

For an arbitrary 2D velocity model, a non-trivial consideration is the determination of ray take-off angle from the source such that an arbitrary receiver location is reached, or such that a particular layer or interface in the model is reached. Today, most 2D ray tracing algorithms offer an automatic determination of take-off angles by shooting trial rays followed by a bisection scheme to sufficiently refine the take-off angles so that the receivers are hit (e.g., Zelt and Smith, 1992). The robustness of the shooting method is crucial because the data from receivers for which it is not possible to find rays will not be used in the inverse step. Also, it is important to be sure that an inability to find rays is because of the velocity model, e.g., a shadow zone, as opposed to a shortcoming of the ray tracing algorithm.

To avoid the limitations of shooting/bisection algorithms, particularly for 3D models, ray bending methods were developed (Um and Thurber, 1987). In this approach, a ray connecting a source and receiver is estimated, typically using a 1D reference model, and the ray's path is iteratively updated using Fermat's principle until the minimum-time path is determined. In practice, shooting methods are more efficient and sufficiently robust for 2D models, whereas bending methods are favored for 3D models.



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 1 Shallow crustal example from the South Portuguese Zone fold-and-thrust belt of first-arrival tomography by Schmelzbach et al. (2008) using the smoothing-regularized Zelt and Barton (1998) algorithm to seek a minimum-structure model. There were 99 shots (grey stars) along the crooked line. (**a**) Velocity model obtained by projecting all the data onto a straight line and performing a 2D inversion. (**b**) Velocity model obtained by performing a 2.5D inversion that preserves the true positions of all shots and receivers. (**c**) Rays projected into the horizontal plane for the 2.5D modeling to illustrate the extreme crookedness of the experiment. Model regions not sampled by rays are grey. The models in (**a**) and (**b**) are very similar, confirming the appropriateness of 2D modeling of traveltime data, even in the case of extreme crooked lines.

In the late 1980s, a new infinite-frequency forward modeling approach was introduced in which first-arrival traveltimes are calculated on a fine grid using a finite-difference solution of the eikonal equation (Vidale, 1988, 1990; Podvin and Lecomte, 1991). Rays are obtained by following the gradient of the time field. These methods are known as wavefront tracking algorithms. Reflected rays can be calculated in a two-step procedure by calculating downgoing and upgoing wavefronts (Hole and Zelt, 1995). Hole and Zelt (1995) also presented a modification to Vidale's approach which otherwise breaks down for velocity contrast of more than ~40%. The advantage of wavefront tracking methods is that they find the ray between any two points with the shortest traveltime, including the diffracted path in the case of a geometrical shadow zone, and they can be very efficient, especially for 3D models. The disadvantages are that they are computationally cumbersome for calculating later arrivals, and they can be inaccurate and/or computationally intensive for high-contrast media. Another type of wavefront tracking scheme that solves the eikonal equation is the fast-marching method, which is computationally efficient and unconditionally stable (Sethian and Popovici, 1999).

Nakanishi and Yamaguchi (1986) and Moser (1991) introduced the shortest path ray-tracing method based on Fermat's principle which uses a fine network of nodes and graph theory. It is capable of handling arbitrarily heterogeneous media and calculating first arrivals, reflections, and multiples. Its advantages over other methods are robustness and the fact that the grid does not have to be rectangular or regular. However, its memory and computation requirements are greater. Van Avendonk et al. (2001) developed a hybrid of the shortest path and ray bending methods that uses the shortest path algorithm to find an initial ray, and then refines it using a bending method. The result is an efficient and robust algorithm to calculate accurate traveltimes and rays for refractions and reflections in 2D and 3D media.

Traveltime tomography and inverison

The traveltime t between a source and receiver along a ray L is given in integral form for a velocity field $v(\mathbf{r})$ as

$$t = \int_L \frac{1}{v(\mathbf{r})} \mathrm{d}l$$

where $\mathbf{r}$ is the position vector in 2D or 3D media. This is a nonlinear inverse problem given the relationship between the measured data (traveltimes) and the unknown model parameters (the velocity field). However, by transforming variables to use slowness $s(\mathbf{r})$, the reciprocal of velocity, instead of velocity as the unknown, a seemingly linear inversion problem is created:

$$t = \int_L s(\mathbf{r}) \mathrm{d}l.$$

However, the ray L is also dependent on the velocity (or slowness) model, thus making the inverse problem nonlinear regardless of what form of model variable or parameterization is used. In global seismology, a 1D (radially-symmetric) earth model is often used to predict the rays given the relatively small velocity heterogeneities in the mantle, and thereby converting the inverse problem into a linear one. In controlled-source seismology, there is no concept of a reference model that is sufficiently accurate to predict the rays in advance given the relatively large velocity heterogeneities that are typically encountered in the crust. This means the controlled-source tomography or inverse problem is always treated as a nonlinear one. This also means the model can be parameterized any number of ways using velocity or slowness, and cells, nodes, or splines, since the problem's nonlinearity must be dealt with regardless of the parameterization. Most often a linearized gradient approach is applied in which a starting model is used and both the model and rays are updated over a series of iterations with the hope that there will be convergence to an acceptable model (the final model).

The model is almost always discretized using cells, nodes, splines, or other interpolating functions; in the latter two cases, the discrete model parameters are the coefficients of the interpolating functions. In the simplest formulation of the tomography problem, the model is parameterized using constant-slowness cells, in which case the equation for the ith data becomes

$$t_i = \sum_j l_{ij} s_j$$

where l_{ij} is the length of the ith ray in the jth model cell and s_j is the slowness in the jth cell. In vector–matrix form this is

$$\mathbf{t} = \mathbf{L}\mathbf{s}.$$

Applying a Taylor series expansion to this equation keeping only the linear term and assuming the rays are independent of slowness yields the perturbation equation

$$\delta\mathbf{t} = \mathbf{G}\delta\mathbf{s}$$

where $\delta\mathbf{t}$ is the data misfit vector equal to the difference between the observed traveltimes and those predicted by a prior model, and $\delta\mathbf{s}$ is the difference between the unknown slowness model and the prior slowness model; the unknown model is also called the estimated model. The partial derivative matrix, $\mathbf{G}$, contains the elements $g_{ij} = \partial t_i / \partial s_j$, and for the constant-slowness cell parameterization, $g_{ij} = l_{ij}$, or $\mathbf{G} = \mathbf{L}$. In the general case, the elements of the partial derivative matrix are $g_{ij} = \partial t_i / \partial m_j$ where m_j is the jth model parameter, which could be the velocity at a node or the coefficient of an interpolating function, or the position of an interface within the velocity model (e.g., Zelt and Smith, 1992). In the general case, the perturbation equation becomes

$$\delta\mathbf{t} = \mathbf{G}\delta\mathbf{m}.$$

The elements of the partial derivative matrix $\mathbf{G}$ are usually calculated analytically to avoid the potential inaccuracy of numerical differencing and the extra computation needed to trace additional rays. The approximations involved in the analytic partial derivatives, e.g., the stationary ray assumption, are typically not a problem since the resultant gradient direction in model space will be improved over a series of iterations.

The perturbation equation is not solved directly because the unknown model parameters are typically under- or mix-determined (Menke, 1989), depending on the model parameterization; with relatively few model parameters, it is possible for the problem to be overdetermined, but even in this case the perturbation equation is not solved directly because it is wise to constrain the magnitude of the model perturbation to avoid violating the linearization assumption. Thus, model constraints, in addition to the observed data, are usually included to stabilize the solution. Also, all observed data contain noise and to avoid over-fitting noisy data, additional model constraints are required to select one model from an infinite number that will statistically predict the observed data at the equivalent desired misfit according to the estimated noise level. This issue points out the non-uniqueness of all inverse problems involving real data, regardless of whether the system of equations is overdetermined or underdetermined.

The additional model constraint is called regularization (e.g., Scales et al., 1990), and in addition to stabilizing a solution, it is an effective way of steering the solution towards models with desirable pre-defined characteristics. The regularization is most often in the form of the zero-, first- or second-order spatial derivatives of the estimated model parameters, or their perturbation from a background model. This is because it is often desirable to seek an estimated model that is as close as possible to another model, and/or a model that has a minimum structure as measured by its spatial derivatives in keeping with Occams's principle (Constable et al., 1987). Regularization is always arbitrary, and the specific form of

regularization that is best for a particular problem depends on the model parameterization, the data, including its coverage and noise level, and the experimental objectives.

The most common forms of regularization in traveltime problems are the minimization of the model's perturbation from a background model, and/or the minimization of the second-order spatial derivative of either the model, or its perturbation from a background model. A model with minimum perturbation is called the smallest model; a model with minimum second-order derivative is called the smoothest model. Any combination of model constraints may be included as a part of the regularization to form the objective function, $\Phi(\mathbf{m})$, where $\mathbf{m}$ is the vector containing the estimated model parameters. The objective function usually measures the square of the data misfit vector and the square of the perturbation or spatial derivative operators comprising the regularization so that minimizing the objective function leads to a linear system of equations that is amenable to efficient algorithms for their exact or approximate solution. One example of an objective function for the traveltime problem is

$$\Phi(\mathbf{m}) = \delta\mathbf{t}^{\mathrm{T}}\mathbf{C}_d^{-1}\delta\mathbf{t} + \lambda\left\{\beta\left[\alpha\left(\mathbf{v}^{\mathrm{T}}\mathbf{W}_h^{\mathrm{T}}\mathbf{W}_h\mathbf{v}+s_z\mathbf{v}^{\mathrm{T}}\mathbf{W}_v^{\mathrm{T}}\mathbf{W}_v\mathbf{v}\right)\right.\right.$$
$$\left.\left.+(1-\alpha)\Delta\mathbf{v}^{\mathrm{T}}\mathbf{W}_p^{\mathrm{T}}\mathbf{W}_p\Delta\mathbf{v}\right]+(1-\beta)\left[\mathbf{z}^{\mathrm{T}}\mathbf{W}_z^{\mathrm{T}}\mathbf{W}_z\mathbf{z}\right]\right\}$$

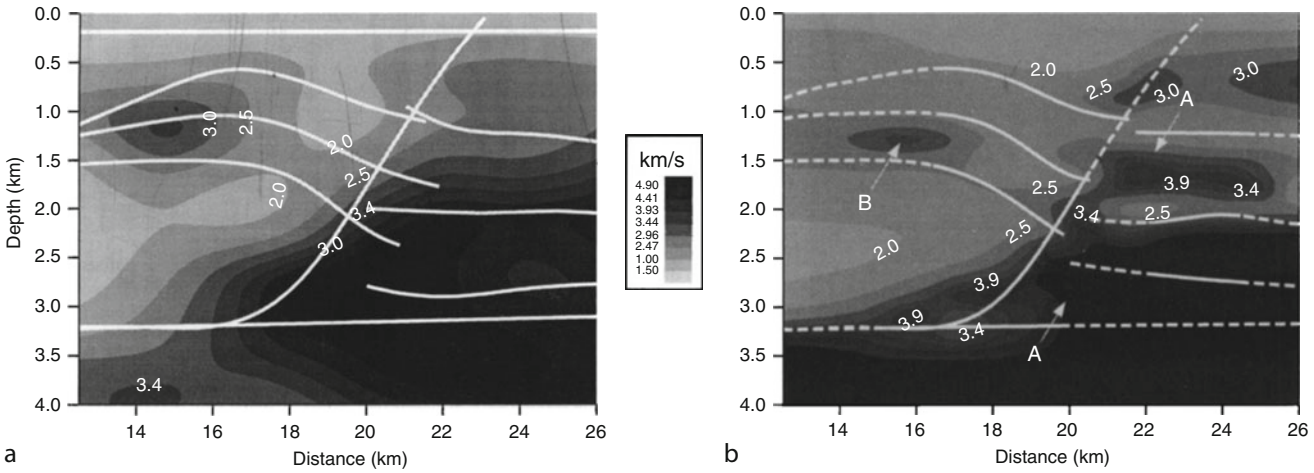
where $\mathbf{m} = \mathbf{v} + \mathbf{z}$; $\mathbf{v}$ and $\mathbf{z}$ are vectors containing the estimated model parameters defining the velocity (or slowness) field and the interfaces (if any); $\Delta\mathbf{v}$ is the velocity perturbation vector equal to $\mathbf{v}$-$\mathbf{v}_o$, and $\mathbf{v}_o$ is the background velocity model vector. $\mathbf{C}_d$ is the data covariance matrix containing the estimated pick uncertainties providing the appropriate up-weighting and down-weighting of the low-noise and high-noise data, respectively. $\mathbf{W}_h$ and $\mathbf{W}_v$ are the horizontal and vertical roughness matrices containing the second-order spatial finite-difference operators that measure the roughness of the velocity field in the horizontal and vertical directions, respectively; $\mathbf{W}_p$ is the velocity perturbation weighting matrix which is a diagonal matrix containing the weights applied to the perturbation between each estimated and background velocity model parameter; $\mathbf{W}_z$ is the interface roughness matrix containing the second-order spatial finite-difference operators that measure the roughness of the interfaces. When applying spatial derivative operators to the velocity field, it is common to use different operators to measure the horizontal and vertical structure, and weight these differently, since one would normally expect the earth to contain more heterogeneity vertically compared to horizontally. Any or all of these operators may be normalized by the reference model values so that relative quantities are penalized as opposed to absolute values (e.g., Toomey et al., 1994). In addition, the operators can be weighted according to their spatial position in the model, e.g., penalize model structure more in the deep portion of the model versus the shallow portion since one would generally expect decreasing resolution with depth.

There are four free parameters in this example objective function to control the relative weight of each term. λ determines the overall amount of regularization, that is, the trade-off between fitting the data and constraining the model; β determines the trade-off between velocity and interface regularization; α determines the trade-off between second derivative and perturbation regularization of the velocity parameters; and s_z specifies the relative weight of vertical versus horizontal smoothing regularization of the velocity parameters. In some algorithms, λ is not a free parameter because it is reduced automatically by the algorithm at each iteration from a free-parameter starting value, λ_o (e.g., Zelt and Barton, 1998). The reduction of λ stabilizes the inversion by constraining the long-wavelength structure in the initial iterations and allowing progressively finer model structure to enter in later iterations.

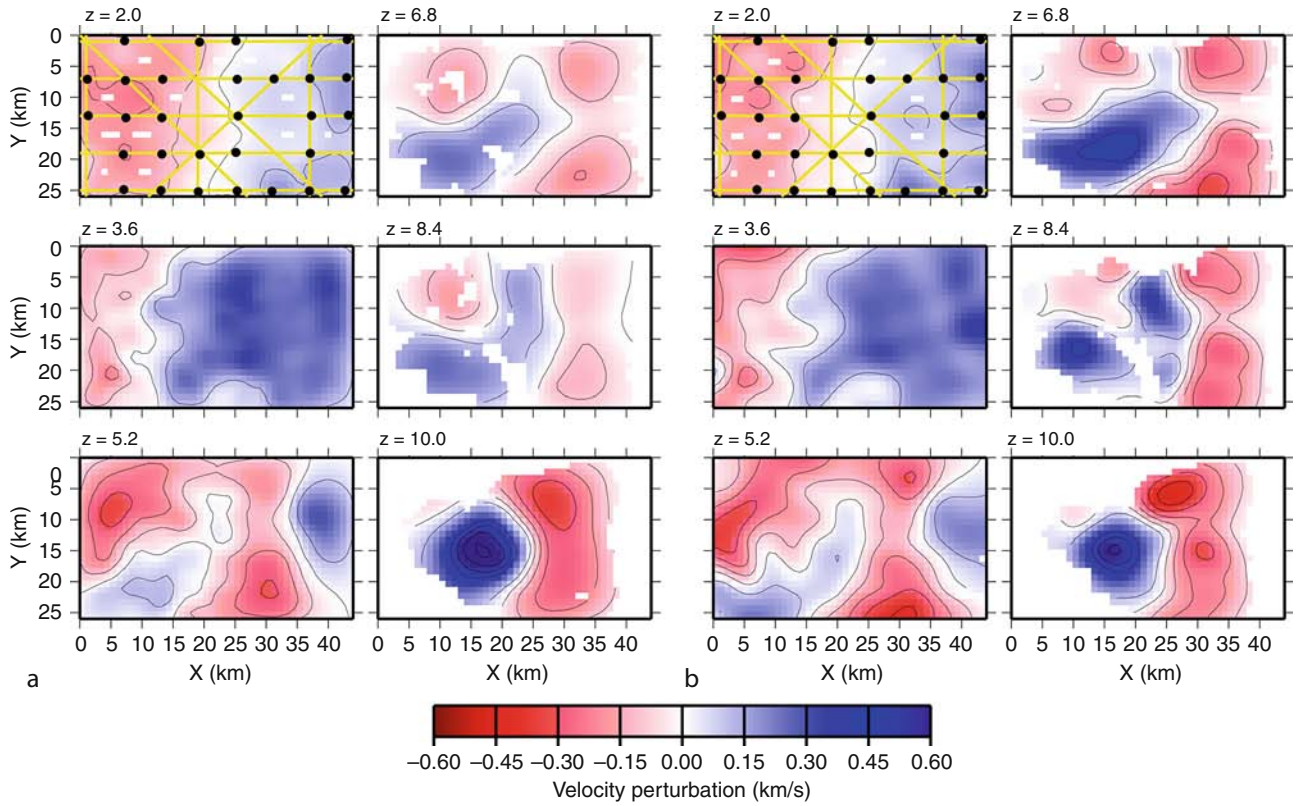
Minimizing the objective function with respect to the unknown model parameters ($\mathbf{v}$ and $\mathbf{z}$) leads to a linear system of equations that may be relatively small or very large, depending on the number of data and model parameters. If the linear system is large, it will also typically be very sparse (at least 99.9%), in which case there are efficient algorithms for their solution (e.g., Paige and Saunders, 1982; Rawlinson et al., 2001). If the linear system is relatively small, it can be solved directly using standard matrix inversion routines, such as LU decomposition (Press et al., 1992).

The objective function presented is designed to illustrate in a single example the different types of regularization that are most commonly applied. In practice, the objective function will typically be simpler depending on the data, model parameterization, and experimental goal. This objective function can be simplified or modified to correspond to those used in most of the popular traveltime inversion and tomography algorithms used today. For example, the Zelt and Smith (1992) inverse method includes only perturbation constraint on the velocity and interface parameters with respect to the model from the previous iteration; the Zelt and Barton (1998) tomographic method includes only smoothness constraints on the velocity field. Thus, together with their very different model parameterizations to be discussed in the next section, these two algorithms can been viewed as end-members of the same regularized inverse approach, and most algorithms in use today fall somewhere in between. In fact, the main differences between the algorithms discussed in the next section lies not in the details of the objective function, but more so in the form of model parameterization, forward calculation, and the types of arrivals considered.

For layered models, including both velocity and interface parameters in a single inversion scheme, as opposed to solving for each parameter separately, either one after the other, or by alternating between the two, makes the



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 2 Synthetic example of reflection tomography for a smooth velocity model by Lailly and Sinoquet (1996) based on real data from a passive margin with a salt structure and a listric fault. (**a**) True model. The reflectors (white curves) are embedded in the smooth velocity field and are allowed to cross each other. (**b**) Estimated model using reflections from the continuous portions of the reflectors (white curves); the parts of the reflectors represented by dashed lines are not illuminated by rays.



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 3 Example of 3D first-arrival tomography applied to ocean-bottom-seismometer (OBS) data from the Faeroe Basin to find the minimum-structure model (Zelt and Barton, 1998). (**a**) Model obtained using smoothing-regularized tomography. (**b**) Model obtained using back-projection tomography method of Hole (1992). Horizontal slices of velocity perturbations with respect to a 1D starting model are shown at the depths labeled from 2 to 10 km. The OBS locations and shot lines are overlaid on the z = 2 km slice. Regions not sampled by rays are blank. Contour interval is 0.1 km/s. Both models provide the same level of fit to the data, but the model from back-projection contains more structure overall. The high-velocity anomaly at z = 10 km likely indicates a basement high.

problem significantly more challenging. There will be more model non-uniqueness because of trade-offs between the two parameter types, and more prior information will be required to deal with the increased ambiguity. Also, more testing will be required to ensure that the trade-offs are fully understood. But solving for both parameter types simultaneously allows one to explore the trade-offs more thoroughly and is the only way to ensure the overall minimum-structure model is obtained, if that is the objective.

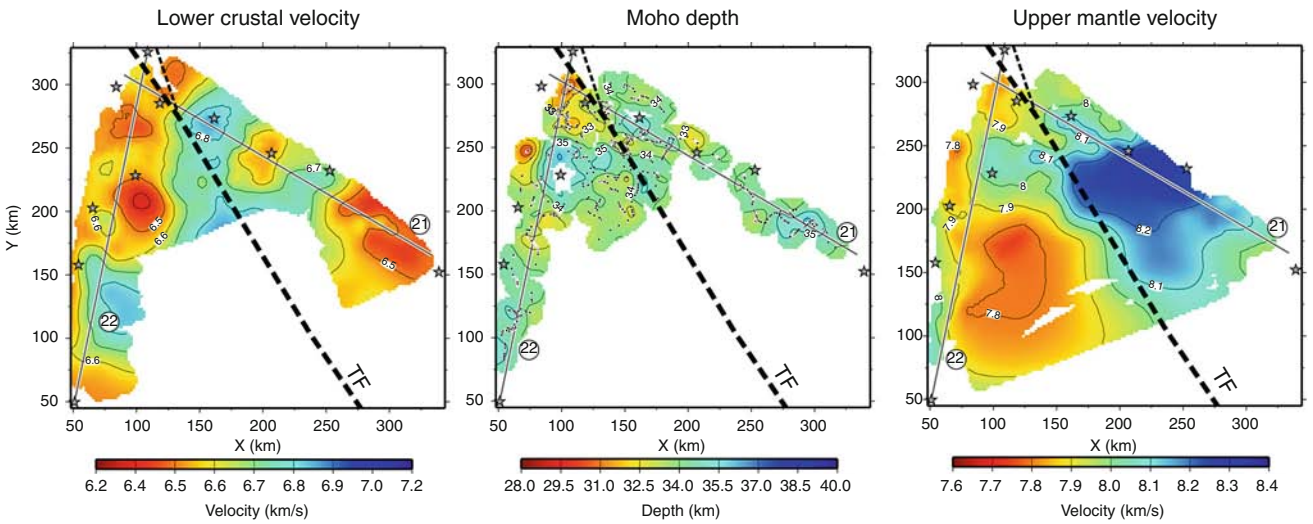
Assuming uncorrelated, Gaussian picking errors, and a model parameterization that does not limit the degree to which the predicted data can match the variability of the observed data, a final model should be chosen that provides a normalized misfit, χ^2, of one (Bevington, 1969).

$$\chi^2 = \frac{1}{N}\left[\delta\mathbf{t}^{\mathrm{T}}\mathbf{C}_d^{-1}\delta\mathbf{t}\right]$$

where N is the number of data. A model that provides a χ^2 value less than one means the observed data are being over-fit, and therefore to some extent, the noise in the data is also being fit. Although the χ^2 criterion is a good rule of thumb, there are several instances in which it should not be followed. First, traveltime picking errors are likely not uncorrelated. Second, for spatially sparse data from a region with strong lateral heterogeneities, it may be necessary to use so many independent model parameters in order to achieve $\chi^2 = 1$ that the constraint on some parameters is unacceptably small, in which case fewer parameters may be preferable and a χ^2 value greater than one is allowed (Zelt, 1999). An inability to achieve $\chi^2 = 1$ may be because of inconsistent picking, for example for reciprocal pairs, in which case the inconsistencies should either be corrected or used as the basis for assigning the pick uncertainties in the first place (e.g., Zelt et al., 2006a).

Algorithms

This section describes many of the algorithms available today, each with its own capabilities and limitations. Some of the earliest 2D traveltime tomography and inversion algorithms were presented by Firbas (1981) and White (1989) which used regular parameterizations and only first arrivals, and Spence et al. (1985) which allowed an irregular model grid and later arrivals. Lutter and Nowack (1990) and Lutter et al. (1990) developed a 2D inversion algorithm using a regular grid of nodes and numerical ray tracing that allows for the independent inversion of first arrivals and reflections for velocity and interface geometry, respectively. The Zelt and Smith (1992) 2D algorithm is the opposite in many respects. An irregular grid of velocity and interface nodes can be used, and any type of refracted or reflected arrival can be inverted simultaneously for velocity and interface geometry. In addition, unique to the Zelt and Smith (1992) algorithm, it allows any subset of the total set of model parameters to be inverted for, holding all others fixed, facilitating an automated forward modeling approach and making it straightforward to incorporate prior information. Clowes et al. (1995) used the Zelt and Smith (1992) algorithm to simultaneously invert refraction and reflection times from

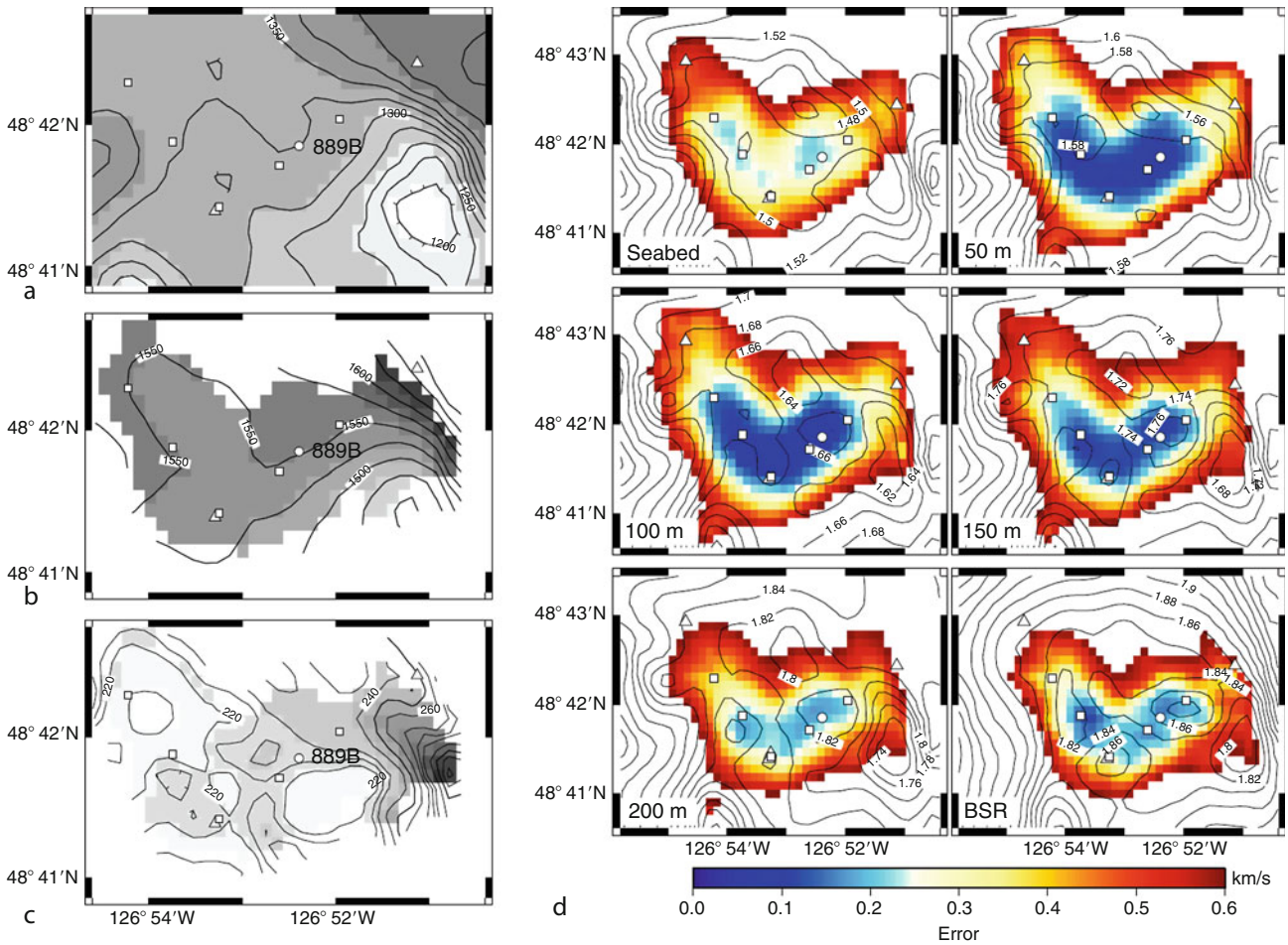


Traveltime Tomography Using Controlled-Source Seismic Data, Figure 4 Example of simultaneous 3D first-arrival and reflection tomography of sparse crustal data across the Tintina strike-slip fault in the northern Canadian Cordillera (Zelt et al., 2006b). Left panel shows the lower crustal velocity averaged vertically between 1 and 15 km above the Moho. Center panel shows depth to Moho; dots show reflection points. Right panel shows the upper mantle velocity averaged vertically between 1 and 15 km below the Moho. Stars are shot locations. The straight lines labeled 21 and 22 indicated the best-fit lines of the crooked profiles containing the receivers locations. Regions not sampled by rays are blank. Dashed line is the surface location of the fault. There is a 0.3–0.4 km/s change in upper mantle velocity across the fault.

a network of 2D profiles such that the model parameters at the intersection points were linked to ensure consistency. This approach can be used to infer 3D structure from 2D data, or develop a starting model for full 3D inversion if offline data is available in addition to inline data (e.g., Zelt et al., 1999).

Hole (1992) developed a 3D first-arrival tomography algorithm using the Vidale (1990) forward modeling scheme and back projection to solve for a smooth velocity field. Hammer et al. (1994) developed a 3D first-arrival tomography algorithm specially suited to sparse data using a spectral, continuous function model parameterization. Toomey et al. (1994) developed a 3D first-arrival tomography algorithm to obtain a smooth velocity field in which the velocity grid is "draped" from an irregular bathymetric surface by vertically shearing the columns of nodes to accurately model seafloor relief. Vertically sheared grids are particularly important for marine data, especially at mid-ocean ridges, where a large velocity discontinuity may exist. Other algorithms with sheared grids have been developed by Van Avendonk et al. (1998) and Korenaga et al. (2000), although they are limited to 2D models, but include reflections. The main difference between these two algorithms is the former uses first arrivals that can turn above or below a reflecting interface at which there is no velocity discontinuity, whereas the latter only uses first arrivals that turn above a reflecting interface.

Lailly and Sinoquet (1996) developed a 2D algorithm for inverting reflection times to estimate a smooth velocity

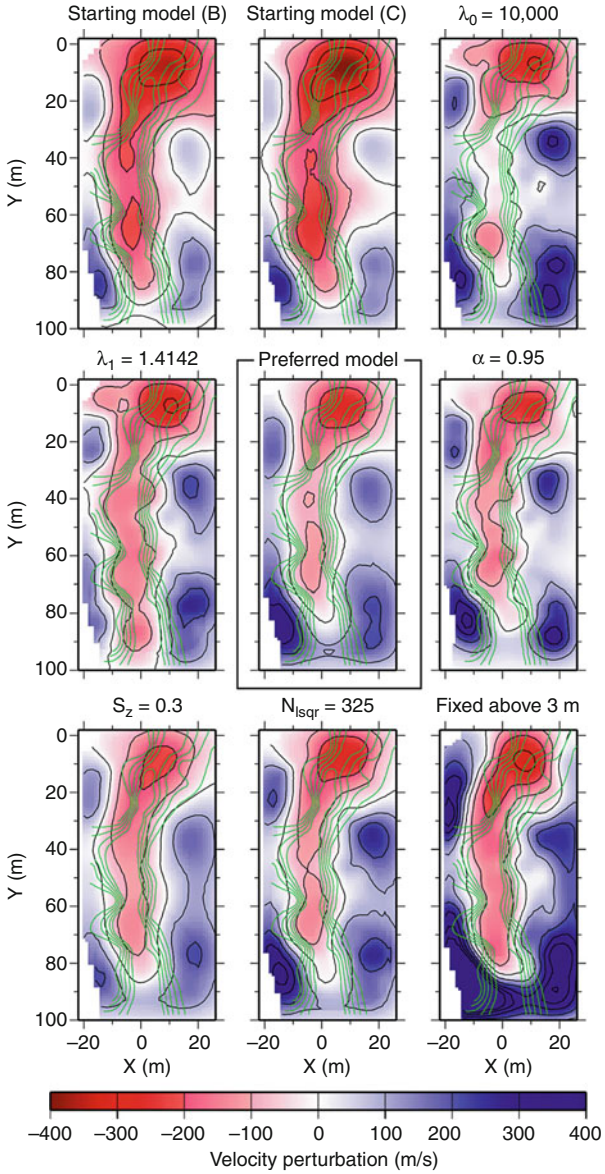


Traveltime Tomography Using Controlled-Source Seismic Data, Figure 5 Example of simultaneous 3D refraction and reflection tomography of OBS data and coincident single channel reflection profiles over a gas hydrate stability zone offshore Vancouver Island (Hobro et al., 2005). (**a**) Seafloor depth. (**b**) Bottom simulating reflector (BSR) depth. (**c**) BSR depth below the seafloor. (**d**) Slices through the velocity model at the seafloor, at 50-m intervals through the hydrate stability zone and immediately above the BSR. Labeled contours indicate velocities (km/s) and color scale marks formal uncertainty estimates. Ten OBS positions and ODP site 889 are marked. Refracted arrivals from above and below the BSR were inverted along with wide-angle and normal-incidence reflections from the BSR.

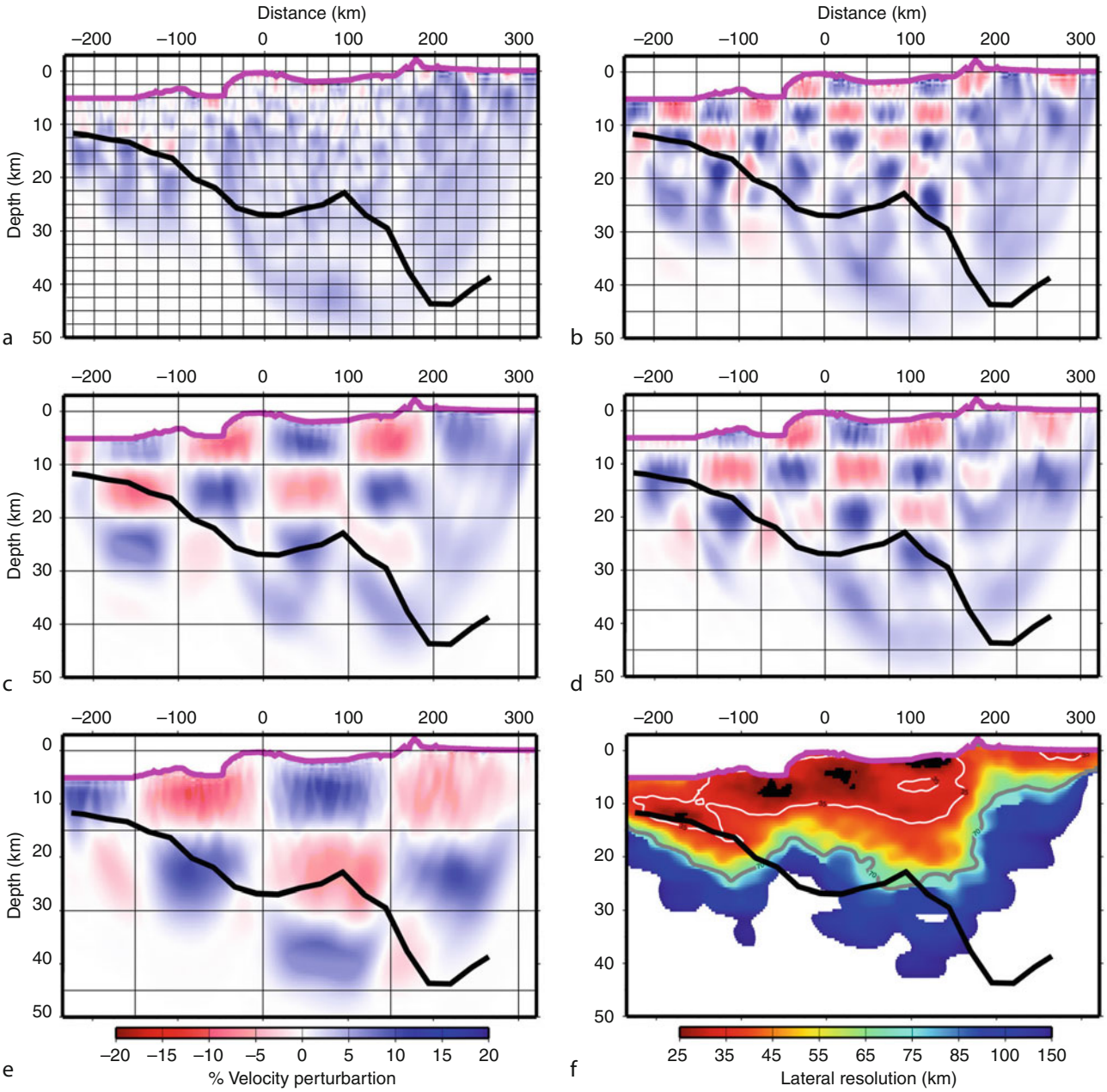
model for pre-stack depth migration of reflection data using the wavefront tracking method of Podvin and Lecomte (1991) for the forward calculation (Figure 2). Zelt and Barton (1998) developed a 3D first-arrival tomography algorithm using smoothing regularization and compared the results with those from the back-projection method of Hole (1992) (Figure 3). The results showed that for the same level of fit to the data, regularized inversion can provide a simpler model. Zelt et al. (2006b) applied 3D simultaneous refraction and reflection tomography to solve for smooth velocities and multiple interfaces using a sparse crustal-scale 3D dataset (Figure 4). McCaughey and Singh (1997) and Hobro et al. (2003) developed related 2D and 3D simultaneous refraction and reflection tomography algorithms that solve for smooth velocities and interfaces with the allowance for discontinuities across the layer boundaries (Figure 5).

Most of the algorithms described above are general purpose in nature and follow more or less from the objective function presented in the previous section. However, more specialized algorithms have been developed and a few are described here to give a sense of what is possible. Hole et al. (1992) developed an algorithm for determining a 3D interface using first arrivals with known velocities above and below the interface. Zhang et al. (1998) inverted traveltime curves instead of points, specifically average slowness and apparent slowness. They claim this balances the contribution from short and long rays, and enhances resolution and convergence speed. Rawlinson and Sambridge (2003b) developed an algorithm for the inversion of refraction and reflection times using a 3D multi-layered model parameterization. Interfaces are defined by a non-uniform node distribution and velocities vary linearly with depth so rays are calculated analytically as piece-wise circular arcs using a shooting method. Trinks et al. (2005) presented a method for simultaneous refraction and reflection inversion using a 2D layered model parameterization which adapts to non-uniform ray coverage such that the cell size is inversely proportional to the local ray density.

Refraction/wide-angle reflection traveltimes can be jointly inverted with coincident near-vertical reflection times in several ways (Zelt, 1999). Arrivals in pre-stack reflection data can be picked and inverted jointly with the wide-angle data, but it is often difficult to pick weak events in pre-stack data. The most common approach is to pick reflections from a stacked reflection section and invert these data simultaneously with the wide-angle data by modeling zero-offset reflections (e.g., McCaughey and Singh, 1997; Zelt et al., 2003). One potential pitfall when using coincident reflection data is incorrectly correlating the near-vertical reflection event with the corresponding wide-angle event or layer boundary; Jaiswal et al. (2006) present a way to avoid this problem. As an alternative to joint inversion, the reflector geometries and interval velocities from stacked data can be used as prior information in the regularization to constrain a layered velocity model (e.g., Bosch et al., 2005).

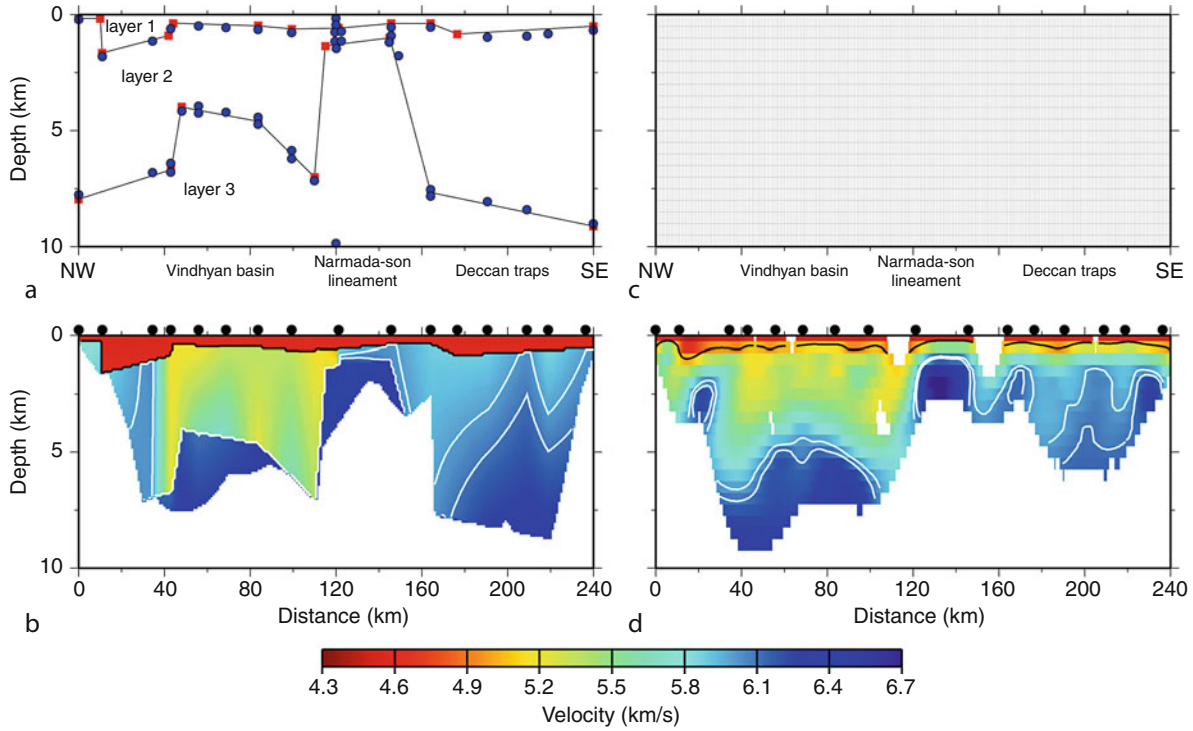


Traveltime Tomography Using Controlled-Source Seismic Data, Figure 6 Example of model assessment for near-surface 3D first-arrival tomography at a groundwater contamination site (Zelt et al., 2006a) using the Zelt and Barton (1998) regularized inversion algorithm seeking a minimum-structure model. Target of the surface was a low-velocity paleochannel cut into a clay layer running roughly north-south through the center of the survey area. The depth to the clay layer was known from extensive well data in the area; depth-to-clay contours (green) from 7 to 11 m overlay the depth slices at 10 m for this comparison of nine different models. Models displayed as perturbations relative to a 1D starting model; the starting velocity at this depth is 1150 ms/s. The preferred final model is in the center. The other models were obtained by trying two different starting models, different values for five free parameters that control the inversion, and by fixing the model above 3 m at the starting model values. Overall, the preferred model has minimum structure; the model in the lower left appears simpler in this depth slice, but it is rougher vertically.



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 7 Example of 2D first-arrival checkerboard test for crustal data from an onshore-offshore experiment across the South American-Caribbean plate boundary at 67°W (Magnani et al., 2009). There were shots and receivers both onshore and offshore. (**a**)–(**e**) show recovered checkerboard anomalies for sizes from 25×2.5 km to 150×15 km. (**f**) Estimated lateral resolution of the velocity model using a 2D version of the method presented by Zelt (1998). Regions with better than 25 km resolution are black; regions with worse than 150 km resolution are white. Pink line indicates the land surface and bathymetry. The coast line is at ~160 km. Black line indicates the Moho from the preferred final model.

The close relationship between seismic velocity and density (Brocher, 2005) lends gravity data to a joint inversion with traveltime data. Nielsen and Jacobsen (2000) used the Zelt and Smith (1992) algorithm as the basis for a simultaneous inversion of refraction and reflection times with gravity data to derive a layered 2D crustal model. Korenaga et al. (2001) carried out a joint inversion of traveltime and gravity data in which error propagation from the velocity model to the predicted gravity anomalies was taken into account.

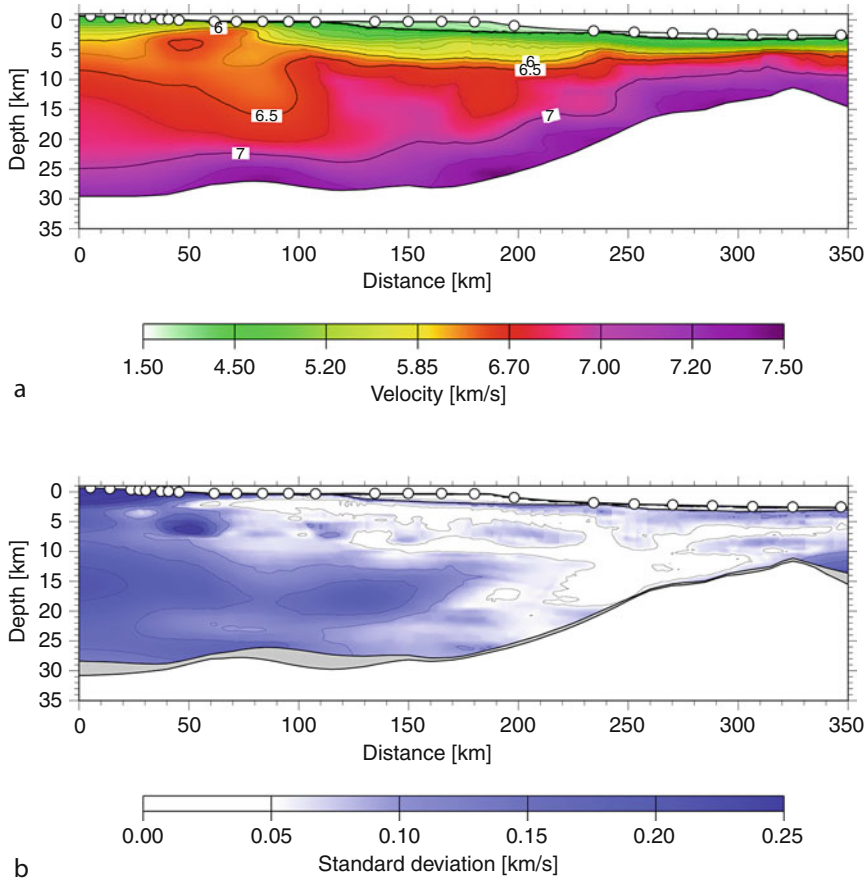


Traveltime Tomography Using Controlled-Source Seismic Data, Figure 8 Example from central India across the Narmada-Son lineament of applying two different traveltime inversion algorithms to the same refraction/wide-angle reflection data (Zelt et al., 2003) to compare the "geological" model with the minimum-structure model. (**a**) Layered model parameterization using the Zelt and Smith (1992) algorithm; there are 43 velocity nodes (blue dots) and 21 interface nodes (red squares). (**b**) Velocity model corresponding to parameterization in (**a**); both first arrivals and reflections from the top of the third layer were used. (**c**) Fine-grid model parameterization using the Zelt and Barton (1998) algorithm to seek a smooth minimum-structure model; there are 7018 slowness cells. (**d**) Velocity model corresponding to parameterization in (**c**); only first arrivals were used. Only those portions of the models sampled by rays are shown. Isovelocity contours of 5.0 km/s (black) and 6.0 and 6.1 km/s (white) indicated. Shot point locations (black dots) and geologic features labeled above the models. An interpretation of the main features of the geological model in (**b**) is supported by the minimum-structure model in (**d**) since it contains the same features.

Model assessment

Once a model to explain a particular set of data is developed, model assessment should be used to explore the non-uniqueness, resolution, and errors associated with the model. This is an attempt to quantify the robustness of the model, a process that is not straightforward for nonlinear inverse problems, and different assessment techniques will be appropriate depending on the data, model parameterization, geologic target, and objectives of the experiment. There are two classes of assessment methods: linear and nonlinear. The former methods assume it is valid to consider only a local neighborhood of the final model within which the model-data relationship is linear. The latter methods involve additional nonlinear inversions of the real data or synthetic data using the true source-receiver geometry of the experiment. Linear methods are quick and easy to apply, although potentially less reliable depending on the nonlinearity of the problem. Nonlinear methods are computational intensive, often more so than what was needed to derive the final model, and may require a specialized inversion algorithm. However, nonlinear methods are the only way to properly estimate the final model's robustness. Zelt (1999) and Rawlinson and Sambridge (2003a) describe many of the assessment techniques in use today, while Zelt et al. (2006a) apply several different techniques to the same 3D dataset (Figure 6).

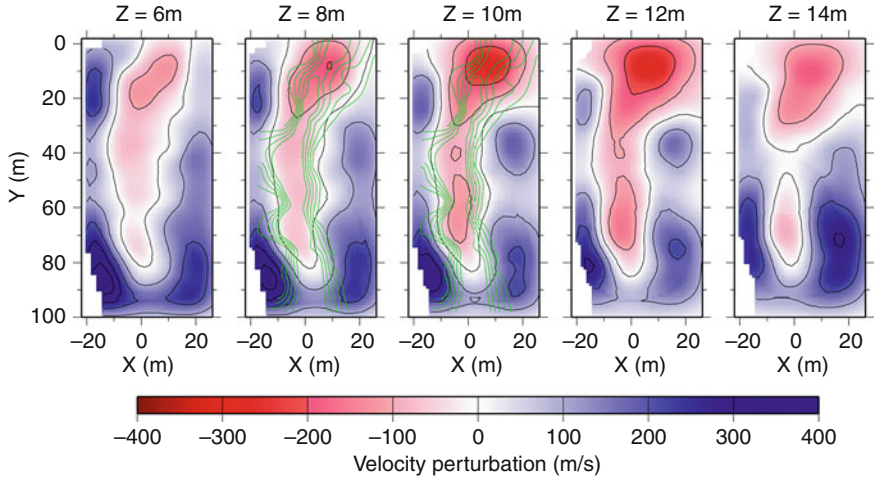
The simplest form of model assessment is to examine the ray coverage in the model using ray plots or "hit counts" (the number of rays sampling each model parameter). These can be misleading since a limited distribution of ray angles may provide less independent model constraint than expected from the number of rays alone. For example, rays in the uppermost mantle will tend to be sub-horizontal, and therefore capable of providing little lateral resolution. Toomey et al. (1994) used a measure called derivative weight sum (DWS) that equals the sum of the partial derivatives for each model parameter. This is somewhat more meaningful than hit count since it weights



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 9 Example of model assessment using a nonlinear Monte Carlo procedure to estimate model variance (Korenaga et al., 2000). (**a**) Velocity model for the southeast Greenland continental margin obtained by joint traveltime inversion of first arrivals and wide-angle Moho reflections recorded at 18 OBS's and eight land stations (white circles). This model was obtained by averaging 100 Monte Carlo ensembles. (**b**) Corresponding standard deviation for velocity and depth nodes; the standard deviation of the Moho is indicated by the width of the grey zone at the base of the crust.

the contribution from each ray according to the parameter's sensitivity to that ray. However, it also fails to account for the degree of independence within the ray set. Other common linear assessment measures include an examination of the diagonals of the posterior resolution and covariance matrices (e.g., Zelt and Smith, 1992). These are more precise measures since they account for the independence of the ray set, but they are best used in a relative as opposed to absolute sense, since they do not account for the nonlinearity of the problem. Rows of the resolution matrix, known as resolution kernels, provide a spatial sense of the averaging of the true structure by the model (Zelt, 1999). Probably the most precise linear assessment method uses singular value decomposition (SVD) since it is able to quantify the nature and degree of model constraint through construction of an orthogonal set of basis vectors that span model space, each with a specified weight in the model reconstruction (e.g., White, 1989; Scales et al., 1990).

A nonlinear assessment of the spatial resolution or uncertainty of a single model parameter, or set of parameters, is possible in which the real data and synthetic data are inverted in the same way that the final model was derived, allowing the full nonlinearity of the problem to be accounted for, including trade-offs between model parameters (Zelt and White, 1995; Zelt, 1999; Christeson et al., 1999). For spatial resolution, the value of a model parameter is perturbed enough to yield a significant traveltime anomaly with respect to the pick uncertainties. Rays are traced through the perturbed model to calculate a set of perturbed traveltimes. The perturbed data are then inverted using the final model as the starting model. The spatial resolution about the selected parameter will be indicated by the amount that the values of adjacent parameters differ from their corresponding value in the original final model. If the model is poorly resolved about the selected parameter, then the parameter's perturbation will be smeared into adjacent parameters, perhaps both



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 10 Example of near-surface 3D first-arrival tomography at a groundwater contamination site (Zelt et al., 2006a) using the Zelt and Barton (1998) regularized inversion seeking a minimum-structure model. The Horizontal depth slices from 6 to 14 m displayed as perturbations relative to a 1D starting model. Black contour interval is 100 m/s. Target of the surface was a low-velocity paleochannel cut into a clay layer running roughly north-south through the center of the survey area. The depth to the clay layer was known from extensive well data in the area; depth-to-clay contours (green) from 7 to 11 m overlay the 8 and 10 m depth slices. This experiment consisted of ~600 shots recorded by ~600 receivers yielding ~360,000 traces, providing 187,877 useable picks for the inversion.

velocities and interface depths if both parameter types are involved, and the extent of the smearing indicates the spatial resolution. Both positive and negative parameter perturbations should be tested, and it will likely be sufficient to examine only one or two representative velocity and interface parameters for each layer or region of the model.

A parameter-selective algorithm such as the Zelt and Smith (1992) approach allows for a nonlinear estimate of a single parameter's absolute uncertainty (Zelt, 1999). The value of the model parameter is slightly perturbed from its value in the final model and held fixed while inverting the observed data involving all other model parameters. The size of the perturbation is increased until the recovered model is unable to fit the observed data as well as the preferred final model based on an F test comparing the traveltime fits of the preferred and perturbed final models. The maximum parameter perturbation that allows a comparable fit to the observed data is an estimate of its absolute uncertainty. Again, both positive and negative parameter perturbations should be tested, and it will likely be sufficient to examine only one or two representative velocity and interface nodes for each layer or region of the model. A fine grid tomographic approach can also be used to perform tests like these. For example, Zelt and Barton (1998) examined one region of a model where there was significant lateral structure even though the constraint from the ray coverage was known to be low in this area. They added regularization to the inversion to force that region to remain laterally homogeneous and thereby establish the required trade-offs elsewhere in the model.

One of the most common forms of nonlinear model assessment to estimate spatial model resolution is the checkerboard test (e.g., Magnani et al., 2009; Figure 7). In these tests an alternating pattern of high and low anomalies is superimposed on the starting model from the inversion of the real data. Synthetic data are calculated for the "checkerboard" model, and then inverted using the same starting model and source-receiver geometry as the real data. The recovered model will closely resemble the checkerboard pattern in regions of good constraint, but will otherwise not resemble the checkerboard model. The resolution at different length scales can be estimated by testing anomaly patterns of different sizes. Anomaly patterns with different polarity, registration, and orientation should be tested to average out the effects of changing ray coverage due to the nonlinearity of the problem (Zelt, 1998).

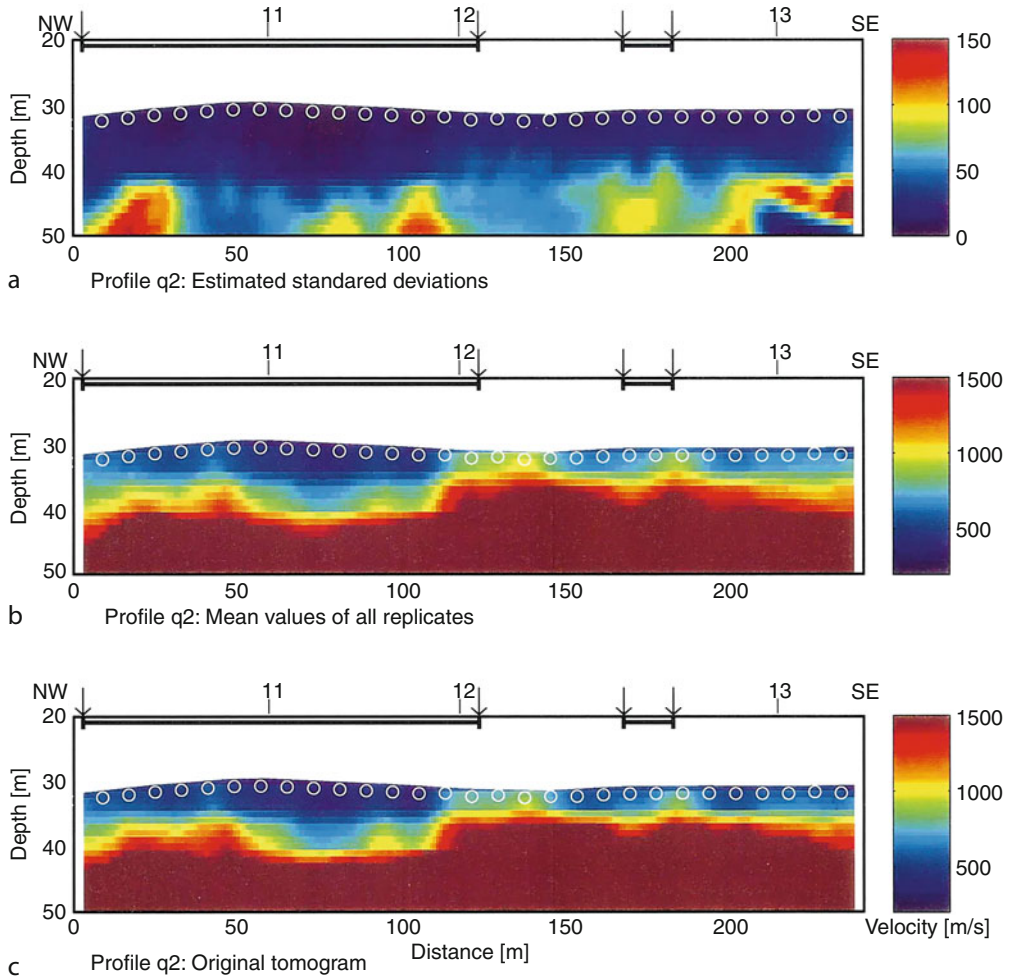
Other nonlinear assessment techniques include trying different starting models, different values of the free parameters in the objective function, and exclusion of subsets of the data considered less reliable (e.g., Zelt et al., 2006a; Figure 6). Using different model parameterizations and different inversion algorithms can be effective, especially when one inversion algorithm seeks a model that satisfies all notions of what is geological reasonable, and one algorithm seeks the minimum-structure model (Figure 8; Zelt et al., 2003). In this way, it is possible to determine what model structure is consistent with the data to facilitate hypothesis testing, while at the same time establishing what model structure is required by the data.

Perhaps the most complete and computationally intensive form of model assessment involves a nonlinear Monte Carlo procedure. Many inversions are performed in which randomized data and/or randomized starting models are used. The resulting models are used to compute posterior model covariance and resolution estimates. Zhang et al. (1998) and Korenaga et al. (2000) applied this approach to 2D inversions of first arrivals and reflections (Figure 9).

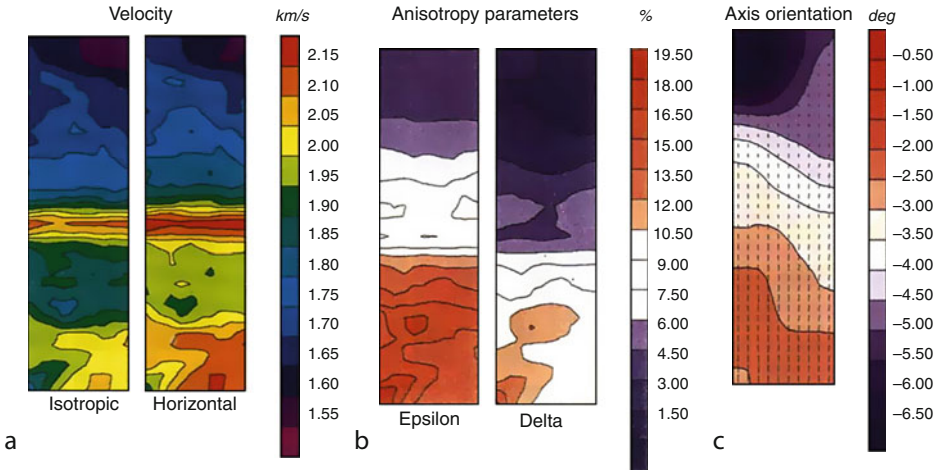
More examples

The examples in Figures 1–9 demonstrate the types of models that can result from a wide range of controlled-source experiments. The remaining examples presented in this section are intended to round out the range of applications to which traveltime tomography and inversion are applied in controlled-source studies, and the types of model structure that can result. The main purpose of the examples is to show what is possible in terms of tomographic imaging of different kinds of geologic features in the near-surface, upper crust, lower crust, and uppermost mantle using traveltime data. All examples shown in this article use only P-wave traveltime data.

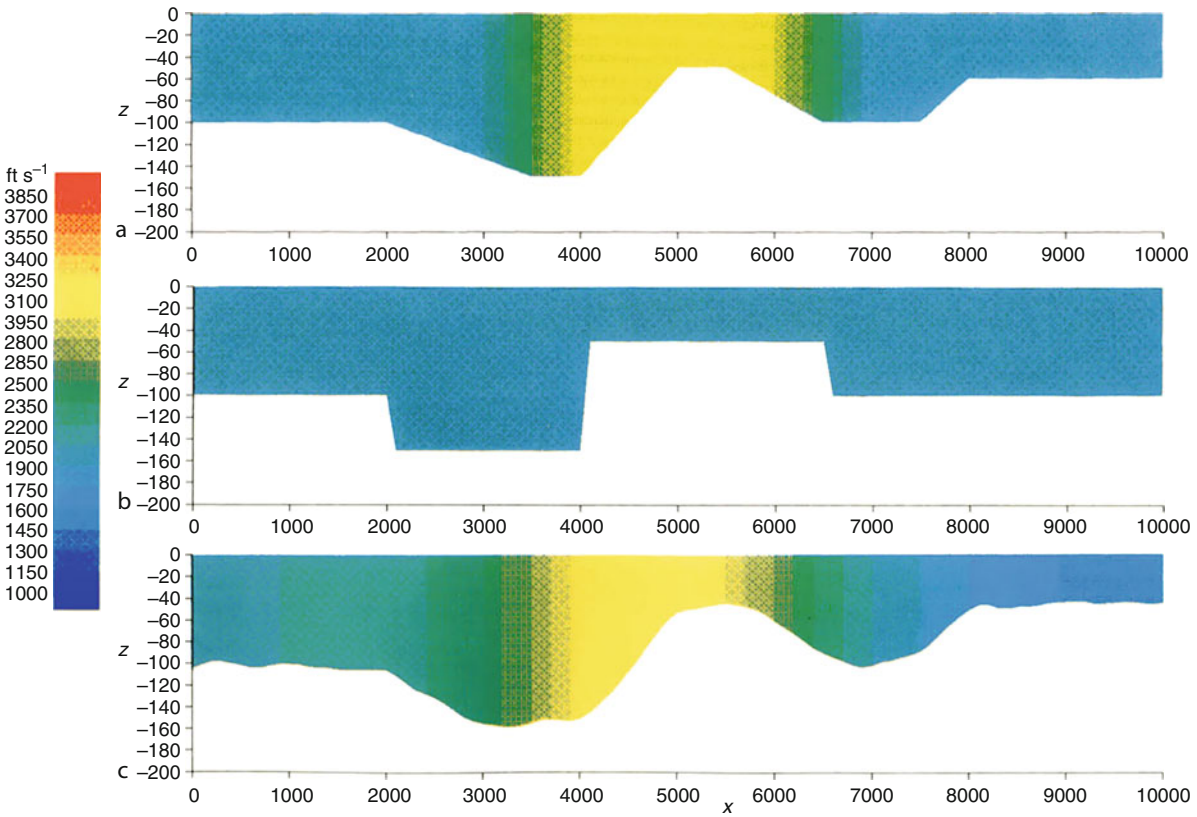
Zelt et al. (2006a) applied the regularized 3D first-arrival tomography algorithm of Zelt and Barton (1998) to a dense 3D dataset from a groundwater contamination site (Figure 10). Lanz et al. (1998) applied 2D first-arrival tomography and a Monte Carlo scheme to estimate a smooth model with standard deviations for a buried waste disposal site (Figure 11). Pratt and Chapman (1992) applied 2D first-arrival tomography to crosswell data to simultaneously determine velocity and anisotropy



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 11 Example of near-surface first-arrival tomography over a buried waste disposal site seeking a smooth model (Lanz et al., 1998). (**a**) Estimated standard deviation of velocities computed using 200 randomized datasets. (**b**) Mean velocity model from 200 randomized datasets. (**c**) Preferred final model from inversion of original picked data. Open circles are shot locations. Arrows indicate landfill borders defined by several other geophysical techniques.



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 12 Example of crosswell tomography in which transverse isotropy is assumed and regularization was used to seek a smooth model (Pratt and Chapman, 1992). The velocity and anisotropic parameters were solved for simultaneously. (**a**) Comparison of isotropic velocity model and horizontal velocity model from anisotropic inversion. The model is 20 m across and 60 m deep. (**b**) Anisotropy parameters. The amount of anisotropy generally increases with depth, reaching nearly 20%. (**c**) Symmetry axis orientation. The length of the markers is proportional to one of the anisotropy parameters (epsilon).

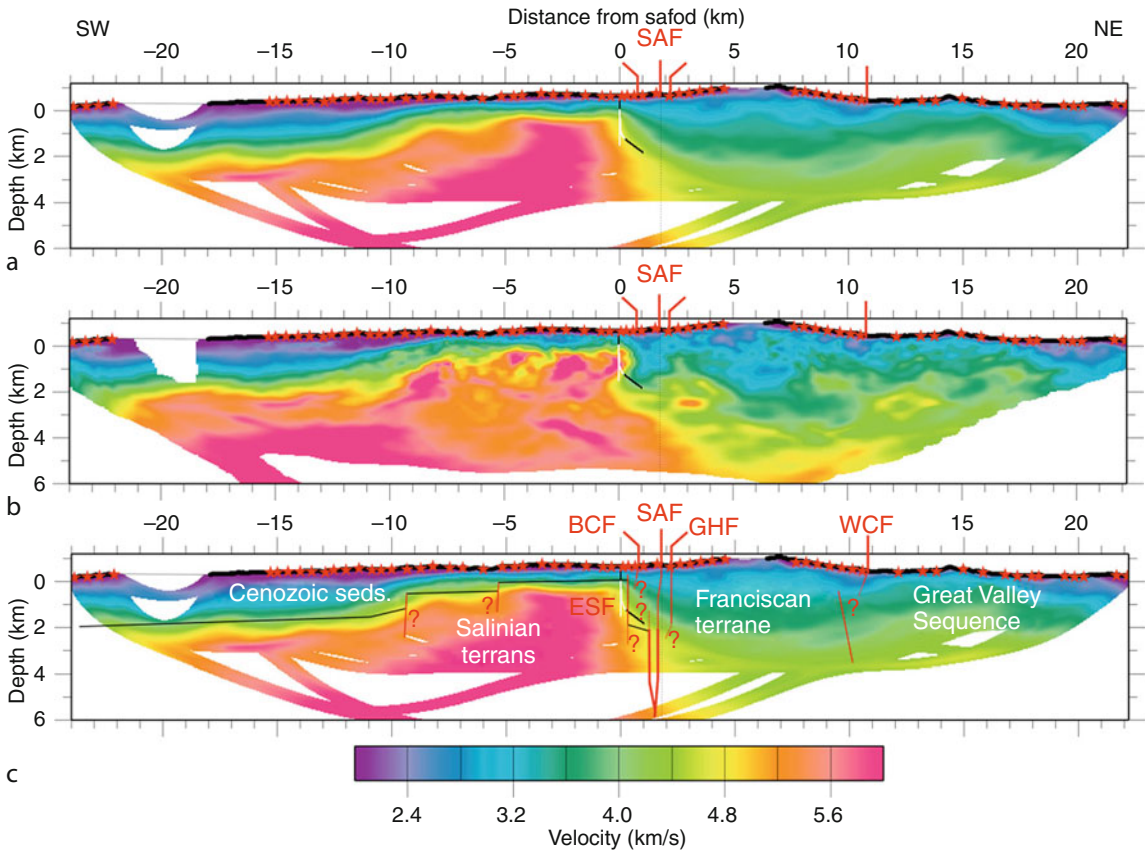


Traveltime Tomography Using Controlled-Source Seismic Data, Figure 13 Example of imaging the near-surface weathering layer (i.e. the refraction statics problem) using smoothing regularization (Scales et al., 1990). Shots were placed at either end of the model and there were 100 receivers between them. (**a**) True model. Refractor velocity is 5000 ft/s. (**b**) Starting model with blocky thickness variations in the weathering layer. Refractor velocity is 4000 ft/s. (**c**) Final model. Refractor velocity is 5001 ft/s. The inversion solved simultaneous for a laterally-varying weathering-layer velocity and refractor, and a constant refractor velocity.

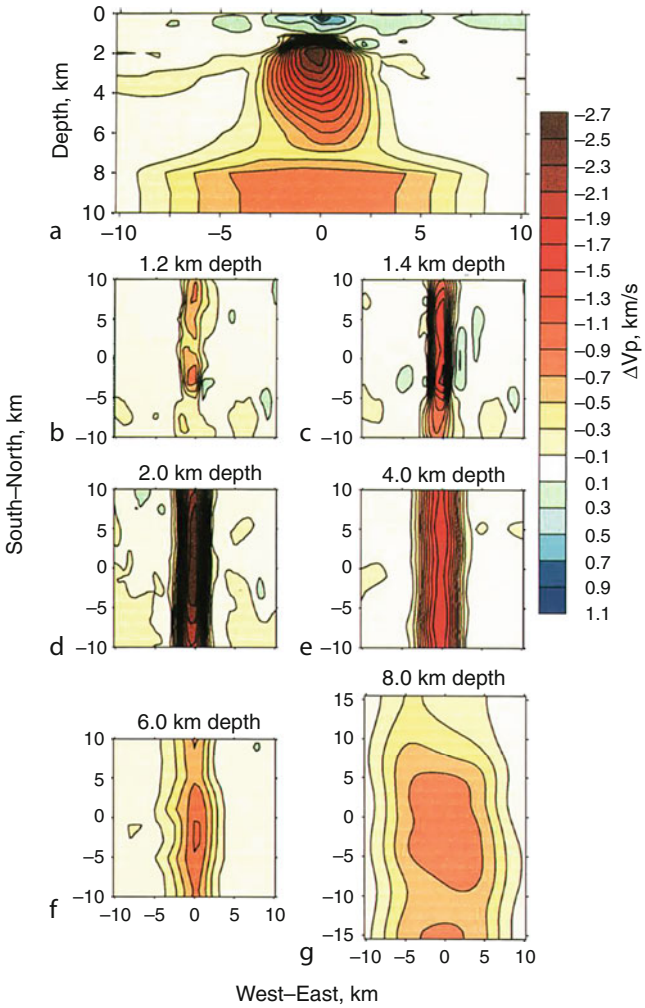
parameters (Figure 12). Scales et al. (1990) imaged a near-surface weathering layer using 2D first-arrival times, solving simultaneously for lateral variations in the velocity and thickness of the layer (Figure 13). Hole et al. (2006) applied 2D first-arrival tomography to data recorded across the San Andreas Fault using two different algorithms (Figure 14). Dunn et al. (2000) applied 3D refraction and reflection tomography incorporating anisotropy to data from the East Pacific Rise (Figure 15). Van Avendonk et al. (1998) inverted refraction and reflection times for a smooth 2D velocity model and an assumed flat Moho across the Clipperton transform fault (Figure 16). Clark et al. (2008) inverted refractions and reflections from an onshore–offshore survey across the South American-Caribbean plate boundary using a smooth tomography approach for the upper crust, and a layer-based inversion for the lower crust, Moho, and upper mantle (Figure 17). Zelt et al. (2003) compared layer-based simultaneous refraction and reflection inversion with independent smooth refraction and reflection tomography for three different 2D crustal datasets (Figure 18).

Future

Full 3D seismic experiments are likely to become more commonplace, necessitating the need for some of the popular 2D tomography and inversion algorithms to be extended to 3D. It is likely that three-component, S-wave, and converted-wave studies will become more widespread, especially for near-surface environmental and engineering studies, necessitating the need for algorithms that can perform coupled P-wave, S-wave, or Poisson's ratio inversions. The desire to incorporate or determine anisotropy will likely become more commonplace in the



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 14 Example of first-arrival tomography across the San Andreas Fault comparing models from two different algorithms (Hole et al., 2006). There were 63 shots (red stars) and 912 receivers (Thick black line). (**a**) Velocity model from application of the Hole (1992) back-projection method. (**b**) Velocity model from application of Zelt and Barton (1998) smoothing-regularized method. Both models provide roughly the same level of fit to the picks. Regions without ray coverage are white. Well-log observations at SAFOD (San Andreas Fault Observatory at Depth) are white for granite and black for sedimentary rock. Note how high velocities (>5 km/s) better match the known position of granite in the well for the model in (**b**) compared to the model (**a**), suggesting that some of the additional structure in the model in (**b**) may be true. (**c**) Geology and fault interpretation overlaid on the model in (**a**). Black lines are top of basement; red lines are faults.

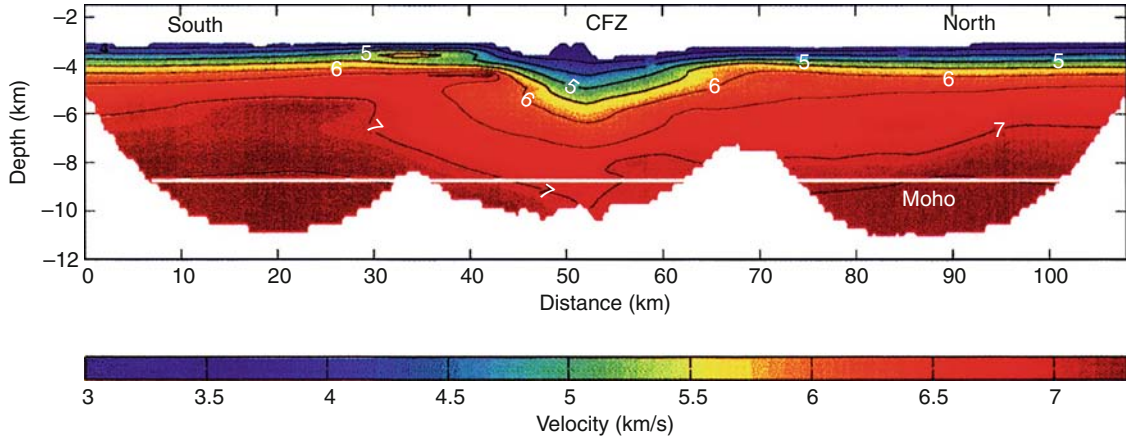


Traveltime Tomography Using Controlled-Source Seismic Data, Figure 15 Example of 3D refraction and reflection tomography across the East pacific Rise at 9°30′N for crust and uppermost mantle structure (Dunn et al., 2000). There were 15 OBS's and 480 explosive sources. Refracted arrivals from the crust and upper mantle and reflections from the Moho were simultaneously inverted. Anisotropy of 4% in the crust and 7% in the upper mantle was included in the modeling. Final velocity model as perturbations relative to a 1D model are shown and are contoured at 0.2 km/s intervals. (**a**) vertical section through the center of the model. (**b**–**g**) Depth slices from 1.2–8.0 km beneath the seafloor. The low-velocity body is interpreted to represent melt distribution within a magmatic system.

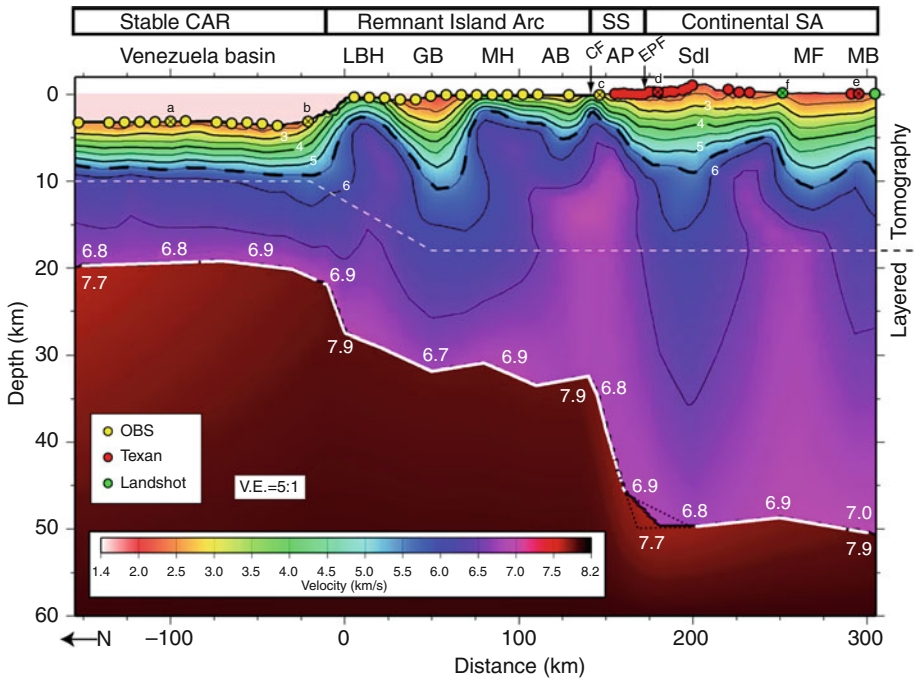
future, and there are currently only a few traveltime algorithms that include any form of anisotropy (e.g., Dunn et al., 2000).

For all scales of seismic experiments, the need for flexible, general-purpose traveltime tomography and inversion algorithms will remain, both as the primary modeling tool and to provide long-wavelength starting models for higher-resolution full waveform inversion techniques (e.g., Brenders and Pratt, 2007). The Zelt and Smith (1992) traveltime inversion algorithm has been widely used for many years for several reasons: (1) an irregular, layered model parameterization adaptable to the data coverage and geologic structures, (2) parameter-selective inversion, (3) any type of body-wave data can be modeled, and (4) it is easy to incorporate prior information. As such it stands apart from most other popular algorithms that use a uniform, fine grid parameterization, and consider only first arrivals, and in some cases reflections. However, the Zelt and Smith (1992) algorithm was primarily intended for relatively sparse 2D data. There will likely be a niche for a new version of the Zelt and Smith algorithm if it maintains all of its current advantages, but is extended to 3D and includes regularization to handle dense data, namely second-order derivatives to seek a smooth, minimum-structure model.

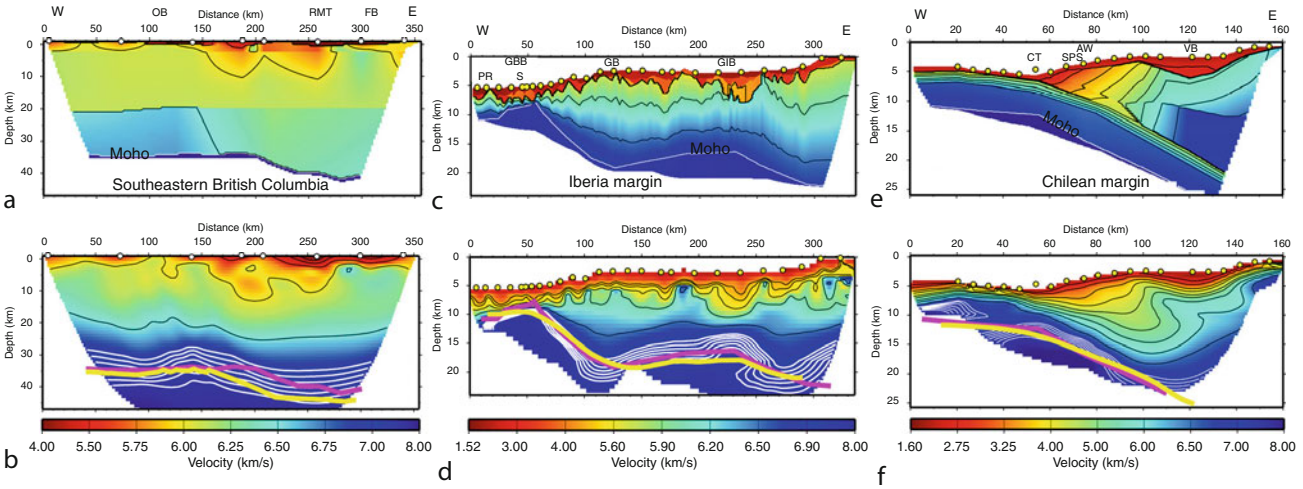
About 10 years ago, a theory for traveltime tomography was developed in global seismology to take into account the finite-frequency nature of seismic data (Dahlen et al., 2000). This method is known as finite-frequency traveltime tomography and it yields what are popularly known as banana-doughnut sensitivity kernels. By taking finite-frequency effects into account, this form of traveltime tomography should, theoretically, yield a more accurate estimation of velocity anomalies in terms of their magnitude and spatial resolution because the physics of wave propagation is treated more accurately. The theory developed for global seismology uses as input data delay times relative to a reference Earth model, and as such the theory is generally not applicable to controlled-source data. This is because for controlled-source data there is no requisite reference velocity model known in advance that is capable of yielding realistic synthetic waveforms that are close enough to the recorded seismograms to yield a meaningful delay time through cross correlation of the waveforms because the crust and near-surface are much more laterally heterogeneous than the mantle. As a result, a nonlinear inverse method must be used, requiring a starting model, an iterative method and recalculated travel paths at each iteration. In addition, the method must be capable of calculating a frequency-dependent traveltime along the total path, as opposed to a delay time. Zelt (2009) presented a finite-frequency traveltime tomography method specifically designed for controlled-source data, that is, a nonlinear inversion of total traveltimes, with frequency taken into account for both the forward calculation of traveltimes, and in the inverse step by calculating the appropriate frequency-dependent sensitivity kernels. The results show it is possible to achieve more accurate velocity estimation than in the equivalent infinite-frequency-derived models. And in addition, applications without any regularization are possible, and therefore finite-frequency traveltime tomography has the potential to allow the data alone to determine the model structure in a robust manner. These characteristics of finite-frequency traveltime tomography



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 16 Example of 2D tomography across the Clipperton transform fault (Van Avendonk et al., 1998). There were only five OBS's. Crust and upper mantle refracted arrivals and Moho reflections were inverted simultaneously for a smooth velocity model. A flat Moho at 8.6 km depth was assumed. The model shows reduced velocities in the upper crust as far as 10 km south and 18 km north of the transform fault (CFZ).



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 17 Example of crustal velocity model from onshore-offshore survey across the South American-Caribbean plate boundary at 64°W (Clark et al., 2008). There were airguns offshore, two shots on land (green dots), OBS's offshore (yellow dots), and seismometers on land (red dots). The model above the dashed white line was obtained by first-arrival tomography using the Zelt and Barton (1998) smoothing-regularized tomography algorithm; the model below the dashed line was derived using the Zelt and Smith (1992) layer-based inversion using upper mantle refractions and Moho reflections to determine a sharp velocity contrast for the Moho. The strike-slip system near the coastline at ~150 km accommodates relative motion between the two plates and from this model is interpreted to extend near vertically through the entire crust and offset the Moho.



Traveltime Tomography Using Controlled-Source Seismic Data, Figure 18 A comparison of three sets of 2D crustal velocity models obtained using two different inversion algorithms (Zelt et al., 2003). The study areas are southeastern British Columbia, the Iberia margin and the Chilean margin. The top row of "geologic" models were obtained using the Zelt and Smith (1992) layer-based algorithm simultaneously inverting crust and upper mantle refractions and Moho reflections. The bottom row of "minimum-structure" models were obtained using the Zelt and Barton (1998) smoothing-regularized first-arrival tomography algorithm. Open circles are shot locations; yellow circles are OBS's. Black contours in the crust have an interval of 0.5 km/s; white contours representing the Moho in the smooth tomographic models have an interval of 0.1 km/s from 7.0–7.6 km/s. Pink lines are the Mohos from the layered models overlain on the tomographic models for comparison. Yellow lines are the Mohos from independent reflection tomography in which the tomographic velocity models were held fixed. Interpretations of the main features of the geological models are supported by the corresponding minimum-structure models since they contain the same features.

suggest that it will play an important role in the future. This is especially true for near-surface studies where infinite-frequency (ray) theory is least likely to be valid because the length scales of heterogeneities are often comparable or smaller than the seismic wavelengths.

Bibliography

Bevington, P. R., 1969. *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill.

Bosch, M., Barton, P., Singh, S. C., and Trinks, I., 2005. Inversion of traveltime data under a statistical model for seismic velocities and layer interfaces. *Geophysics*, **70**, R33–R43.

Brenders, A. J., and Pratt, R. G., 2007. Full waveform-tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, **168**, 133–151.

Brocher, T. M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin. Seismological Society of America*, **95**, 2081–2092.

Cerveny, V., Molotkov, I., and Psencik, I., 1977. *Ray Method in Seismology*. Prague: University of Karlova.

Chambers, G., Levander, A., Zelt, C. A., and Dugan, B., 2009. Seismic waveform tomography with multicomponent data at a groundwater contamination site. *SEG Expanded Abstracts*, **28**, 3994.

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.

Christeson, G. L., McIntosh, K. D., Shipley, T. H., Flueh, E. R., and Goedde, H., 1999. Structure of the Costa Rica convergent margin, offshore Nicoya Peninsula. *Journal of Geophysical Research*, **104**, 25,443–25,468.

Clark, S. A., Zelt, C. A., Magnani, M. B., and Levander, A., 2008. Characterizing the Caribbean-South American plate boundary at 64 W using wide-angle seismic data. *Journal of Geophysical Research*, **113**, B07401, doi:10.1029/2007JB005329.

Clowes, R. M., Zelt, C. A., Amor, J. R., and Ellis, R. M., 1995. Lithospheric structure in the southern Canadian Cordillera from a network of seismic refraction lines. *Canadian Journal of Earth Sciences*, **32**, 1485–1513.

Constable, S. C., Parker, R. L., and Constable, C. G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, **52**, 289–300.

Dahlen, F. A., Hung, S. H., and Nolet, G., 2000. Frechet kernels for finite-frequency traveltimes – I. Theory. *Geophysical Journal International*, **141**, 157–174.

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°30'N. *Journal of Geophysical Research*, **105**, 23,537–23,555.

Dziewonski, A. M., and Anderson, D. L., 1984. Seismic tomography of the earths interior. *American Scientist*, **72**, 483–494.

Firbas, P., 1981. Inversion of travel-time data for laterally heterogeneous velocity structure – linearization approach. *Geophysical Journal of the Royal Astronomical Society*, **67**, 189–198.

Hammer, P. T. C., Dorman, L. M., Hildebrand, J. A., and Cornuelle, B. D., 1994. Jasper Seamount structure: seafloor seismic refraction tomography. *Journal of Geophysical Research*, **99**, 6731–6752.

Hobro, J. W. D., Singh, S. C., and Minshull, T. A., 2003. Three-dimensional tomographic inversion of combined reflection and

refraction seismic traveltime data. *Geophysical Journal International*, **152**, 79–93.

Hobro, J. W. D., Minshull, T. A., Singh, S. C., and Chand, S., 2005. A three-dimensional seismic tomographic study of the gas hydrate stability zone, offshore Vancouver Island. *Journal of Geophysical Research*, **110**, B09102, doi:10.1029/2004JB003477.

Hole, J. A., 1992. Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research*, **97**, 6553–6562.

Hole, J. A., and Zelt, B. C., 1995. Three-dimensional finite-difference reflection travel times. *Geophysical Journal International*, **121**, 427–434.

Hole, J. A., Clowes, R. M., and Ellis, R. M., 1992. Interface inversion using broadside seismic refraction data and 3-dimensional travel time calculations. *Journal of Geophysical Research*, **97**, 3417–3429.

Hole, J. A., Ryberg, T., Fuis, G. S., Bleibinhaus, F., and Sharma, A. K., 2006. Structure of the San Andreas fault zone at SAFOD from a seismic refraction survey. *Geophysical Research Letters*, **33**, L07312, doi:10.1029/2005GL025194.

Jaiswal, P., Zelt, C. A., and Pecher, I. A., 2006. Seismic characterization of a gas hydrate system in the Gulf of Mexico using wide-aperture data. *Geophysical Journal International*, **165**, 108–120.

Korenaga, J., Holbrook, W. S., Kent, G. M., Kelemen, P. B., Detrick, R. S., Larsen, H. C., Hopper, J. R., and Dahl-Jensen, T., 2000. Crustal structure of the southeast Greenland margin from joint refraction and reflection seismic tomography. *Journal of Geophysical Research*, **105**, 21,591–21,614.

Korenaga, J., Holbrook, W. S., Detrick, R. S., and Kelemen, P. B., 2001. Gravity anomalies and crustal structure at the southeast Greenland margin. *Journal of Geophysical Research*, **106**, 8853–8870.

Lailly, P., and Sinoquet, D., 1996. Smooth velocity models in reflection tomography for imaging complex geological structures. *Geophysical Journal International*, **124**, 349–362.

Lanz, E., Maurer, H., and Green, A. G., 1998. Refraction tomography over a buried waste disposal site. *Geophysics*, **63**, 1414–1433.

Levander, A., Zelt, C. A., and Symes, W. W., 2007. Active source studies of crust and lithospheric structure. In Romanowicz, B., and Dziewonski, A., (eds.), *Treatise on Geophysics, Seismology and Structure of the Earth*. Amsterdam, Boston: Elsevier, Vol. 1, pp. 247–288.

Lutter, W. J., and Nowack, R. L., 1990. Inversion for crustal structure using reflections from the PASSCAL Ouachita experiment. *Journal of Geophysical Research*, **95**, 4633–4646.

Lutter, W. J., Nowack, R. L., and Braile, L. W., 1990. Seismic imaging of upper crustal structure using travel times from the PASSCAL Ouachita experiment. *Journal of Geophysical Research*, **95**, 4621–4631.

Magnani, M. B., Zelt, C. A., Levander, A., and Schmitz, M., 2009. Crustal structure of the South American-Caribbean plate boundary at 67°W from controlled source seismic data. *Journal of Geophysical Research*, **114**, B02312, doi:10.1029/2008JB005817.

McCaughey, M., and Singh, S. C., 1997. Simultaneous velocity and interface tomography of normal-incidence and wide-aperture traveltime data. *Geophysical Journal International*, **131**, 87–99.

McMechan, G. A., and Mooney, W. D., 1980. Asymptotic ray theory and synthetic seismograms for laterally varying structures: theory and application to the Imperial Valley, California. *Bulletin. Seismological Society of America*, **70**, 2021–2035.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory.* San Diego: Academic.

Moser, T. J., 1991. Shortest path calculation of seismic rays. *Geophysics*, **56**, 59–67.

Nakanishi, I., and Yamaguchi, K., 1986. A numerical experiment on nonlinear image reconstruction from first-arrival times for two-dimensional island arc structure. *Journal of Physics of the Earth*, **34**, 195–201.

Nielsen, L., and Jacobsen, B. H., 2000. Integrated gravity and wide-angle seismic inversion for 2-D crustal modelling. *Geophysical Journal International*, **140**, 222–232.

Nowack, R. L., and Braile, L. W., 1993. Refraction and wide-angle reflection tomography: theory and results. In Iyer, H. M., and Hirahara, K. (eds.), *Seismic Tomography: Theory and Practice*. London: Chapman and Hall, pp. 733–765.

Paige, C. C., and Saunders, M. A., 1982. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, **8**, 43–71.

Podvin, P., and Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools. *Geophysical Journal International*, **105**, 271–284.

Pratt, R. G., and Chapman, C. H., 1992. Traveltime tomography in anisotropic media–II. Application. *Geophysical Journal International*, **109**, 20–37.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 1992. *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd edn. Cambridge: Cambridge University Press.

Ramachandran, K., Dosso, S. E., Spence, G. D., Hyndman, R. D., and Brocher, T. M., 2005. Forearc structure beneath southwestern British Columbia: a three-dimensional tomographic velocity model. *Journal of Geophysical Research*, **110**, B02303, doi:10.1029/2004JB003258.

Rawlinson, N., and Sambridge, M., 2003a. Seismic traveltime tomography of the crust and lithosphere. *Advances in Geophysics*, **46**, 181–198.

Rawlinson, N., and Sambridge, M., 2003b. Irregular interface parametrization in 3-D wide-angle seismic traveltime tomography. *Geophysical Journal International*, **155**, 79–92.

Rawlinson, N., Houseman, G. A., and Sambridge, M., 2001. Inversion of seismic refraction and wide-angle reflection traveltimes for 3-D layered crustal structure. *Geophysical Journal International*, **145**, 381–401.

Scales, J. A., Docherty, P., and Gersztenkorn, A., 1990. Regularisation of nonlinear inverse problems: imaging the near-surface weathering layer. *Inverse Problems*, **6**, 115–131.

Schmelzbach, C., Zelt, C. A., Juhlin, C., and Carbonell, R., 2008. P- and Sv-velocity structure of the South Portuguese Zone fold-and-thrust belt, SW Iberia, from traveltime tomography. *Geophysical Journal International*, **175**, 689–712.

Sethian, J. A., and Popovici, A. M., 1999. 3-D traveltime computation using the fast marching method. *Geophysics*, **64**, 516–523.

Spence, G. D., Whittall, K. P., and Clowes, R. M., 1984. Practical synthetic seismograms for laterally varying media calculated by asymptotic ray theory. *Bulletin. Seismological Society of America*, **74**, 1209–1223.

Spence, G. D., Clowes, R. M., and Ellis, R. M., 1985. Seismic structure across the active subduction zone of western Canada. *Journal of Geophysical Research*, **90**, 6754–6772.

Toomey, D. R., Solomon, S. C., and Purdy, G. M., 1994. Tomographic imaging of the shallow crustal structure of the East Pacific Rise at 9°30'N. *Journal of Geophysical Research*, **99**, 24,135–24,157.

Trinks, I., Singh, S. C., Chapman, C. H., Barton, P. J., Bosch, M., and Cherrett, A., 2005. Adaptive traveltime tomography of densely sampled seismic data. *Geophysical Journal International*, **160**, 925–938.

Um, J., and Thurber, C., 1987. A fast algorithm for two-point seismic ray tracing. *Bulletin. Seismological Society of America*, **77**, 972–986.

Van Avendonk, H. J. A., Harding, A. J., Orcutt, J. A., and McClain, J. S., 1998. A two-dimensional tomographic study of the Clipperton transform, fault. *Journal of Geophysical Research*, **103**, 17,885–17,899.
Van Avendonk, H. J. A., Harding, A. J., Orcutt, J. A., and Holbrook, W. S., 2001. Hybrid shortest path and ray bending method for traveltime and raypath calculations. *Geophysics*, **66**, 648–653.
Van Avendonk, H. J. A., Shillington, D. J., Holbrook, W. S., and Hornbach, M. J., 2004. Inferring crustal structure in the Aleutian arc from a sparse wide-angle seismic data set. *Geochemistry, Geophysics, Geosystems*, **5**, Q08008, doi:10.1029/2003GC000664.
Vidale, J. E., 1988. Finite-difference calculation of traveltimes. *Bulletin. Seismological Society of America*, **78**, 2062–2076.
Vidale, J. E., 1990. Finite-difference calculation of traveltimes in three dimensions. *Geophysics*, **55**, 521–526.
White, D. J., 1989. Two-dimensional seismic refraction tomography. *Geophysical Journal International*, **97**, 223–245.
Zelt, C. A., 1998. Lateral velocity resolution from three-dimensional seismic refraction data. *Geophysical Journal International*, **135**, 1101–1112.
Zelt, C. A., 1999. Modeling strategies and model assessment for wide-angle seismic traveltime data. *Geophysical Journal International*, **139**, 183–204.
Zelt, C.A., 2009. Frequency-dependent traveltime tomography for controlled-source, near-surface seismic data. EOS Trans. AGU, 90(52), Fall Meet. Suppl., Abstract NS31A–1155.
Zelt, C. A., and Barton, P. J., 1998. Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe Basin. *Journal of Geophysical Research*, **103**, 7187–7210.
Zelt, C. A., and Ellis, R. M., 1988. Practical and efficient ray tracing in two-dimensional media for rapid traveltime and amplitude forward modeling. *Canadian Journal of Exploration Geophysics*, **24**, 16–31.
Zelt, C. A., and Smith, R. B., 1992. Seismic traveltime inversion for 2-D crustal velocity structure. *Geophysical Journal International*, **108**, 16–34.
Zelt, C. A., and White, D. J., 1995. Crustal structure and tectonics of the southeastern Canadian Cordillera. *Journal of Geophysical Research*, **100**, 24,255–24,273.
Zelt, C. A., Hojka, A. M., Flueh, E. R., and McIntosh, K. D., 1999. 3D simultaneous seismic refraction and reflection tomography of wide-angle data from the central Chilean margin. *Geophysical Research Letters*, **26**, 2577–2580.
Zelt, C. A., Sain, K., Naumenko, J. V., and Sawyer, D. S., 2003. Assessment of crustal velocity models using seismic refraction and reflection tomography. *Geophysical Journal International*, **153**, 609–626.
Zelt, C. A., Azaria, A., and Levander, A., 2006a. 3D seismic refraction traveltime tomography at a groundwater contamination site. *Geophysics*, **71**, H67–H78.
Zelt, C. A., Ellis, R. M., and Zelt, B. C., 2006b. 3-D structure across the Tintina strike-slip fault, northern Canadian Cordillera, from seismic refraction and reflection tomography. *Geophysical Journal International*, **167**, 1292–1308.
Zhang, J., ten Brink, U. S., and Toksöz, M. N., 1998. Nonlinear refraction and reflection traveltime tomography. *Journal of Geophysical Research*, **103**, 29,743–29,757.

Cross-references

Deep Seismic Reflection and Refraction Profiling
Earth's Structure, Continental Crust
Ocean Bottom Seismics
Seismic Anisotropy
Seismic Structure at Mid-Ocean Ridges
Seismic, Ray Theory
Seismic, Waveform Modeling and Tomography

TSUNAMI

Steven N. Ward
Institute of Geophysics and Planetary Physics, University of California at Santa Cruz, Santa Cruz, CA, USA

Definition

Dispersive. Characteristic of waves whose velocity of propagation depends on wave frequency. The shape of a dispersive wave packet changes as it moves along.
Eigenfunction. Functional shape of the horizontal and vertical components of wave motion versus depth in the ocean for a specific wave frequency.
Geometrical spreading. Process of amplitude reduction resulting from the progressive expansion of a wave from its source.
Run-up. Final phase of tsunami life starting when the wave shoals to a size equal to the water depth and begins to break, and ending when the water runs over land and reaches its highest level.
Shoal. Process of waves coming into shallow water. Shoaling waves slow, shorten their wavelength, and grow in size.
Wavenumber. Wavenumber k equals 2π divided by wavelength λ. Large wavenumbers associate with short waves and small wavenumbers associate with long waves.

Tsunami are gravity-driven water waves. They belong to the same family as common sea waves that we see every day; however, tsunami are distinct in their mode of generation and in their physical traits. Unlike common sea waves that evolve from persistent winds, most tsunami spring from sudden shifts of the ocean floor. These sudden shifts can originate from undersea landslides and volcanoes, but mostly, submarine earthquakes parent tsunami. Compared to wind-driven waves, tsunami waves have periods, velocities, and wavelengths ten or a hundred times larger and present profoundly different shoreline consequences than do their common cousins.

Tsunami = killer wave?

In the years since the 2004 Sumatra earthquake, everyone has seen disturbing videos of tsunami-caused destruction. Certainly the prospect of a "killer wave" born from some far off earthquake is frightening. More so is the thought that after traveling with stealth great distances, that wave might suddenly rise up without warning at your own doorstep. Could it be me desperately floating by in the next tsunami video?

Understandably, worst-case scenarios of natural hazards come to mind, but it is important to keep perspective.

Tsunami over 2 m high are not common. It takes a submarine earthquake greater than magnitude M8 to source a wave of this size. On a global average, only one M8+ earthquake occurs per year. Of these, maybe one in five strikes under the ocean with an orientation favorable for tsunami excitation. True, tsunami decay in transit and have shorter run-up at distant shores, so wave damage tends to localize within 1,000 km of that one in five quakes. On these accounts, tsunami that induce widespread damage and casualties number only about one per decade. Even at that modest recurrence rate, your personal risk is mitigated increasingly through education and technology. Today, ocean bottom pressure sensors detect stealthy tsunami of a few centimeters height in the open sea. Moreover, with advances in understanding, communication, and implementation since the 2004 Sumatra earthquake, scientists are better, quicker, and more specific at prediction than they were just a decade past. There is a good chance now that you will be warned and take action even if that rare wave does come your way.

Tsunami characteristics

Tsunami period, velocity, and wavelength

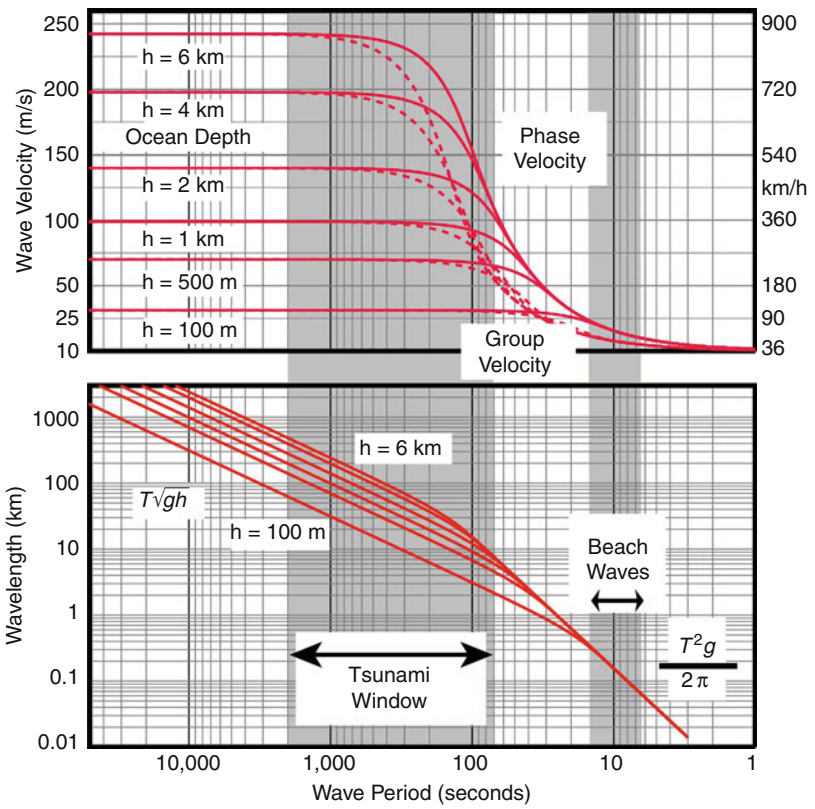
When discussing tsunami, I like to contrast them with something that we all have experienced – waves at the beach. Tsunami belong to the same family as these ordinary ocean waves but with major distinctions: principally tsunami period, tsunami velocity, and tsunami wavelength.

Tsunami period. The period (T) of a wave equals the time elapsed between one passing crest and the next. Common beach waves have a period of about 10 s (darkened column right side Figure 1). The period of tsunami depends upon what creates it (earthquake, landslide, asteroid impact), but I can tell you that tsunami have far greater periods than beach waves. The "tsunami window" in Figure 1 covers waves of 70, 200, 500, or even 2,000 s periods. Unlike "splash and dash" beach waves, tsunami arrive and may continue to flow in for several minutes. I like the description – "The ocean turns into a river." Anyone who has seen those 2004 Sumatra videos can testify that tsunami act more flood-like than wave-like.

Tsunami velocity. Under classical theory, the phase c (ω), and group u(ω) velocity of surface gravity waves on a flat ocean of uniform depth h are:

$$c(\omega) = \sqrt{\frac{gh\ \tanh[k(\omega)h]}{k(\omega)h}} \tag{1}$$

and



Tsunami, Figure 1 Phase velocity c(ω) (*solid lines*) and group velocity u(ω) (*dashed lines*) of tsunami waves on a flat earth covered by oceans of 100 m to 6 km depth (*top*). Wavelength associated with each wave period. The "tsunami window" is marked (*bottom*).

$$u(\omega) = c(\omega)\left[\frac{1}{2} + \frac{k(\omega)h}{\sinh[2k(\omega)h]}\right]. \quad (2)$$

Here, g is the acceleration of gravity (9.8 m/s^2) and $k(\omega)$ is the *wavenumber* associated with a sea wave of frequency $\omega = 2\pi/T$. Wavenumber connects to wavelength as $\lambda(\omega) = 2\pi/k(\omega)$ and to phase velocity as $c(\omega) = \omega/k(\omega)$. Wavenumber also satisfies the relation:

$$\omega^2 = gk(\omega)\tanh[k(\omega)h]. \quad (3)$$

For surface gravity waves spanning 1–50,000 s period, Figure 1 (*top*) plots $c(\omega)$ and $u(\omega)$. These velocities vary widely and increase for waves of longer periods in deeper water. The speed of ordinary sea waves is about 15 m/s (50 km/h) – the speed of a moped. Because of their longer period, waves in the tsunami window travel much more rapidly, reaching 160–250 m/s (600–900 km/h) in the open ocean. It is said that deep-water tsunami travel at the speed of a jet airliner. While true, be aware that wave energy is the quantity that runs ahead at jet speeds. The physical bits of water in a deep-water tsunami move less than a meter per second (see next section). Waves whose velocity varies with frequency are called *dispersive*. During propagation, dispersion "pulls apart" originally pulse-like waves into their component frequencies. Dispersion is strongest for waves whose period falls on the steepest slopes of the group velocity curves in Figure 1.

Tsunami wavelength. Wavelength measures the distance between one wave crest and the next. Wavelength equals the product of wave period times phase velocity $\lambda(\omega) = Tc(\omega)$. Common beach waves have a wavelength of about 100 m – the length of a football field (Figure 1, *bottom*). Tsunami, with their longer period and higher velocity, have much longer wavelengths than beach waves. In the deep ocean, tsunami span 10, 30, even 100 km between crests. If you could stand on one tsunami crest, the next one might be over the horizon. With wavelengths this large, tsunami slopes are very small even if the wave has large amplitude. For ships at sea, tsunami pass completely unnoticed.

Short wave versus long wave: Discussions of waves of length λ in oceans of depth h sometimes include two simplifications: a long wave approximation ($\lambda >> h$, $1/k >> h$) and a short wave approximation ($\lambda << h$, $1/k << h$). Under a long wave approximation [$kh \to 0$, $\tanh(kh) \to$ kh, $\sinh(2kh) \to 2kh$] Equations 1–3 predict nondispersive wave propagation with $c(\omega) = u(\omega) = \sqrt{gh}$. Long wave theory holds for the flat part of the curves in Figure 1 (*top*). Under a short wave approximation, [$kh \to \infty$, $\tanh(kh) \to 1$, $\sinh(kh) \to \infty$] the equations predict dispersive propagation with $c(\omega) = 2u(\omega) = gT/2\pi$. Short wave theory holds to the right in Figure 1 (*top*) where all the curves lie atop each other. Waves in the tsunami window have intermediate character, behaving like shallow water waves at their longest periods and like deep-water waves at their shortest periods. Neither the long nor the short wave simplification serves adequately in tsunami studies. A rigorous treatment requires an approach that works for waves of all lengths.

Tsunami eigenfunctions

Tsunami also differ from waves at the beach in the way they move the ocean. Tsunami *eigenfunctions* describe wave motion in a tsunami mode of a particular frequency. Consider coordinate system (x, y, z) where $\hat{\mathbf{x}}$ points north, $\hat{\mathbf{y}}$ east, and $\hat{\mathbf{z}}$ down. Vertical (u_z) and horizontal (u_x) components of tsunami eigenfunctions normalized to vertical amplitude A_z at the sea surface are:

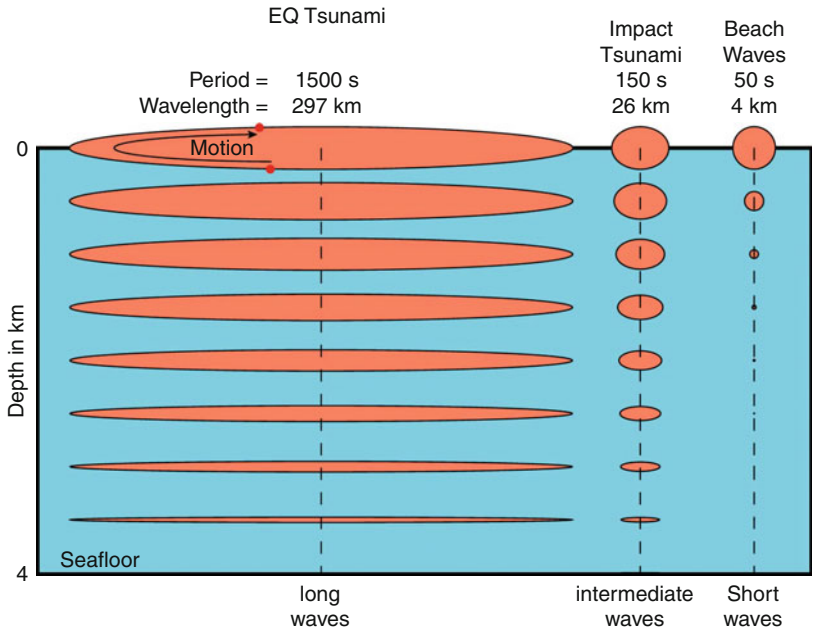
$$u_z(\omega, z) = A_z \frac{k(\omega)g}{\omega^2} \frac{\sinh[k(\omega)(h-z)]}{\cosh[k(\omega)h]} e^{i[k(\omega)x - \omega t]} \quad (4)$$

$$u_x(\omega, z) = A_z \frac{-ik(\omega)g}{\omega^2} \frac{\cosh[k(\omega)(h-z)]}{\cosh[k(\omega)h]} e^{i[k(\omega)x - \omega t]}$$

Figure 2 plots tsunami eigenfunctions versus depth in a 4-km deep ocean at long (1,500 s), intermediate (150 s), and short (50 s) periods. The ellipses trace the path of a water particle as a wave of frequency ω passes. At 50-s period (Figure 2, *right*), ordinary ocean waves have deep-water behavior. Water particles move in circles that decay exponentially from the surface. Sensibly, because the eigenfunctions of short waves do not reach to the seafloor, their velocity is independent of ocean depth [$c(\omega) = 2u(\omega) = gT/2\pi$, Figure 1, *top right*]. The failure of short waves to "feel" the seafloor also means that they cannot be excited by deformations of it. The only means to excite ordinary ocean waves is to disturb the surface.

At 1,500 s period (Figure 2, *left*), the tsunami has a wavelength of λ = 297 km and it acts like a long wave. Although vertical displacement peaks at the ocean surface and drops to zero at the seafloor, horizontal displacement persists undiminished through the ocean column. Unlike ordinary waves that are confined near the sea surface, the energy of a tsunami spreads through the entire depth of the sea. You cannot out dive a tsunami. This fact explains why tsunami can be detected by pressure sensors on the seafloor. A second distinction of tsunami versus ordinary sea waves is that their horizontal motion far exceeds their vertical motion. Every meter of up and down in a 1,500-s tsunami involves ≈10 m of back and forth. If you were to build a shore-based or space-based "tsunami detector," the large horizontal motions might make a better target than the vertical motions. Applications of Doppler radar come to mind.

Toward the short period side of the tsunami window at 150 s (Figure 2, *middle*), λ decreases to 26 km. For these tsunami waves, long wave characteristics begin to break down, and horizontal and vertical motions more closely agree in amplitude.



Tsunami, Figure 2 Tsunami eigenfunctions in a 4-km deep ocean at periods 1,500, 150, and 50 s. Vertical displacements at the ocean surface has been normalized to A_z = 1 m in each case.

From (4), it is easy to deduce the peak velocity of surface water in waves of amplitude A_z as:

$$\left|v_x^{\max}(\omega, z=0)\right| = \frac{A_z k(\omega) g}{\omega} = \frac{A_z g}{c(\omega)}. \tag{5}$$

For long waves $c(\omega)$ equals$\sqrt{gh}$, and peak surface water velocity in a 1-m amplitude tsunami in 4,000 m of waver would be just 4.9 cm/s. Peak water velocity at the sea floor

$$\begin{aligned}\left|v_x^{\max}(\omega, z=h)\right| &= \frac{A_z\, k(\omega) g}{\omega \cosh[k(\omega)h]} \\ &= \frac{A_z g}{c(\omega)\cosh[k(\omega)h]}\end{aligned} \tag{6}$$

is even less. As mentioned, although tsunami travel at several hundred meters per second, the water itself moves at a tiny fraction of this.

Tsunami excitation

Tsunami get started in many ways. Suppose that the seafloor at positions $\mathbf{r}_0$ uplift instantaneously by an amount $u_z^{\text{bot}}(\mathbf{r}_0)$ at time $\tau(\mathbf{r}_0)$. Under classical tsunami theory in a uniform ocean of depth h, this sea bottom disturbance produces surface tsunami waveforms (vertical component) at observation point $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$ and time t of

$$u_z^{\text{surf}}(\mathbf{r}, t) = \text{Re} \int_{\mathbf{k}} d\mathbf{k}\, \frac{e^{i[\mathbf{k}\cdot\mathbf{r} - \omega(k)t]}}{4\pi^2 \cosh(kh)} F(\mathbf{k}). \tag{7a}$$

with

$$F(\mathbf{k}) = \int_{\mathbf{r}_0} d\mathbf{r}_0\, u_z^{\text{bot}}(\mathbf{r}_0) e^{-i[\mathbf{k}\cdot\mathbf{r}_0 - \omega(k)\tau(\mathbf{r}_0)]} \tag{7b}$$

$k = |\mathbf{k}|$, and $\omega^2(k) = gk\tanh(kh)$. The integrals in (7a, 7b) cover all wavenumber space and locations $\mathbf{r}_0$ where the seafloor disturbance $u_z^{\text{bot}}(\mathbf{r}_0) \neq 0$.

Equation 7a looks scary but it has three identifiable pieces:

(a) The $F(\mathbf{k})$ term is the wavenumber spectrum of the seafloor uplift. This number relates to the amplitude, spatial and time distribution of the uplift. Tsunami trains are dominated by wavenumbers in the span where $F(\mathbf{k})$ is the greatest. The peak of $F(\mathbf{k})$ corresponds to the characteristic dimension of the uplift. Large-dimensioned uplifts produce longer wavelength, lower frequency tsunami than small-dimensioned sources.

(b) The $1/\cosh(kh)$ term comes from the tsunami eigenfunction shapes (4) and it acts to low-pass filter the source spectrum $F(\mathbf{k})$. Because $1/\cosh(kh) \to 1$ when $kh \to 0$, and $1/\cosh(kh) \to 0$ when $kh \to \infty$, the filter favors long waves. Due to the low-pass filter effect of the ocean layer, only wavelengths of the uplift that exceed three times the ocean depth (i.e., $kh = 2\pi h/\lambda <\approx 2$) contribute much to tsunami.

(c) The exponential term in (7a) contains all of the propagation information including travel time, *geometrical spreading*, and frequency dispersion.

By rearranging Equation 7a, 7b, vertical tsunami motions at **r** can also be written as:

$$u_z^{\text{surf}}(\mathbf{r},t) = \text{Re}\int_0^\infty \frac{k\,\text{d}k e^{-i\omega(k)t}}{2\pi\cosh(kh)} \sum_{n=-\infty}^{\infty} \text{J}_n(kr)e^{in\theta} F_n(k)$$

with (8)

$$F_n(k) = \int_{\mathbf{r}_0} \text{d}\mathbf{r}_0\, u_z^{\text{bot}}(\mathbf{r}_0)\, J_n(k\text{r}_0)\, e^{i(\omega(k)\tau(\mathbf{r}_0)-n\theta_0)}$$

Here θ marks azimuth from north (the $\hat{\mathbf{x}}$ direction) of the observation point **r** from the coordinate origin. The $J_n(x)$ are cylindrical Bessel functions. For simply distributed uplift sources, (8) might be easier to evaluate than (7a, 7b).

Tsunami excitation by earthquakes

Earthquakes produce 80–90% of tsunami. Not surprisingly, earthquake features determine many sea wave characteristics. Earthquakes result from slip on faults and three primary parameters describe the process – moment, mechanism, and depth.

Moment measures earthquake strength. Moment M_o is the product of rigidity μ of the source region's rocks, fault area A, and average fault slip Δu. Earthquake moment and earthquake magnitude tie through a number of empirical formulae. One formula defines moment magnitude M_w as $M_w = (2/3)(\log M_o - 9.05)$. Earthquake moment varies by 2×10^4 within the magnitude range $6.5 \le M_w \le 9.5$ (Table 1). Even without a detailed understanding of tsunami generation, it is safe to suppose that the larger the earthquake moment, the larger the tsunami, all else fixed.

Mechanism specifies the orientation of the earthquake fault and the direction of slip on it. Usually, faults are idealized as plane rectangles with normal $\hat{\mathbf{n}}$. Three angles then, summarize earthquake mechanisms – the strike and dip of the fault and the angle of slip vector $\hat{\mathbf{a}}$ measured from the horizontal in the plane of the fault. (Seismologists call this angle the "rake.") The role of fault mechanism on tsunami production is not as obvious as the influence of moment; however, one might suspect that earthquakes that affect large vertical displacements of the seafloor would be more effective than faults that make large horizontal displacements.

Earthquake *depth* needs no explanation. Because seafloor shifts cause tsunami, the distance of the fault from the seafloor should be important. Presumably, deep earthquakes produce less potent tsunami than similar shallow earthquakes.

Numerical, or synthetic waveforms quantify the roles of earthquake parameters on tsunami generation. For illustration, insert into (7a, 7b) the surface uplift pattern $u_z^{\text{bot}}(\mathbf{r}_0)$ of a small earthquake fault (point source really) placed at depth d in a halfspace. Further assume that the uplift occurs instantly with $\tau(\mathbf{r}_0) = 0$. (Actually, real earthquakes uplift the seafloor over several, or several tens of, seconds. This distinction is not a big issue because tsunami waves have periods of many hundreds of seconds. Uplifts taking a few dozen seconds to develop look "instantaneous" to tsunamis.) Equation 7a, 7b becomes

$$u_z^{\text{surf}}(\mathbf{r},t) = \int_0^\infty k\,\text{d}k \frac{\cos\omega(k)t}{2\pi\cosh(kh)}\left[A\Delta u \mathbf{M}_{ij}\varepsilon_{ij}\right], \qquad (9)$$

where

$$\begin{aligned}
\varepsilon_{xx} &= -\frac{1}{4}\left(\frac{\mu}{\lambda+\mu} - kd\right)[J_0(kr) - J_2(kr)\cos 2\theta]e^{-kd}\\
\varepsilon_{yy} &= -\frac{1}{4}\left(\frac{\mu}{\lambda+\mu} - kd\right)[J_0(kr) + J_2(kr)\cos 2\theta]e^{-kd}\\
\varepsilon_{xy} = \varepsilon_{yx} &= \frac{1}{4}\left(\frac{\mu}{\lambda+\mu} - kd\right)[J_2(kr)\sin 2\theta]e^{-kd}\\
\varepsilon_{zz} &= -\frac{1}{2}\left(\frac{\mu}{\lambda+\mu} + kd\right)[J_0(kr)]e^{-kd}\\
\varepsilon_{xz} = \varepsilon_{xz} &= \frac{kd}{2}[J_1(kr)\cos\theta]e^{-kd}\\
\varepsilon_{yz} = \varepsilon_{zy} &= \frac{kd}{2}[\text{J}_1(kr)\sin\theta]e^{-kd}
\end{aligned} \qquad (10)$$

The six elements of symmetric tensor

$$M_{jk} = \left(\hat{a}_j\hat{n}_k + \hat{n}_j\hat{a}_k\right) \qquad (11)$$

capsulize the mechanism of the earthquake. In (11), $\hat{\mathbf{n}}$, $\hat{\mathbf{a}}$ are the fault normal and slip vectors introduced above. A pure dip slip earthquake on a vertical north–south

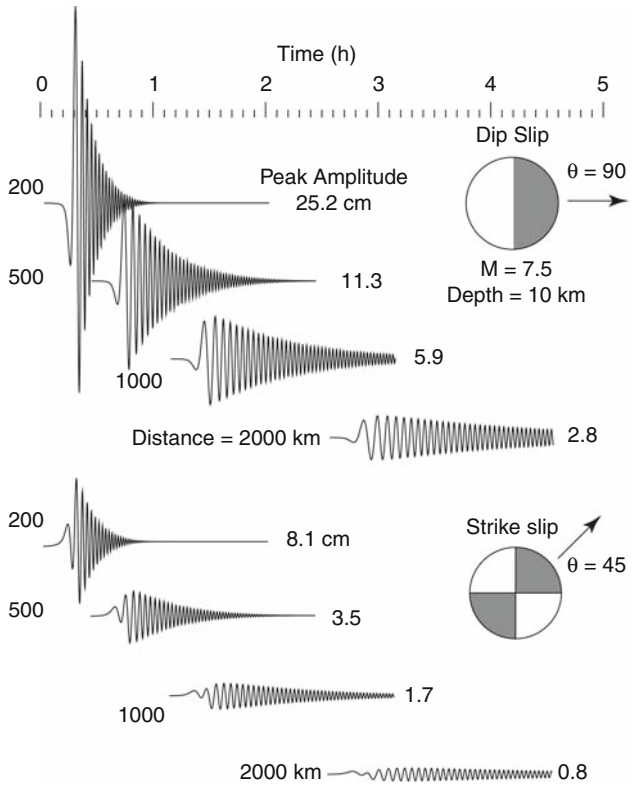
Tsunami, Table 1 Relationship between earthquake magnitude and moment with values of fault area, length, and mean slip for typical tsunami-generating earthquakes. This paper assumes $\log(L) = 0.5\,M_w - 1.8$, $\Delta u = 2 \times 10^{-5}\,L$, and $\lambda = \mu = 5 \times 10^{10}$ Pa

Magnitude M_w	Moment M_0 (Nm)	Area A (km^2)	Length L (km)	Width W (km)	Slip Δu (m)
6.5	6.3×10^{18}	224	28	8	0.56
7.0	3.5×10^{19}	708	50	14	1.00
7.5	2.0×10^{20}	2,239	89	25	1.78
8.0	1.1×10^{21}	7,079	158	45	3.17
8.5	6.3×10^{21}	22,387	282	79	5.66
9.0	3.5×10^{22}	70,794	501	141	10.0
9.5	2.0×10^{23}	223,872	891	251	17.8

trending fault for instance, has $\hat{\mathbf{n}} = \hat{\mathbf{y}}$ and $\hat{\mathbf{a}} = \hat{\mathbf{z}}$, so $M_{yz} = M_{zy} = 1$ and $M_{xx} = M_{yy} = M_{zz} = M_{xz} = M_{zx} = M_{xy} = M_{yx} = 0$.

The bracketed terms in (9) contain all of the relationships between earthquake parameters and tsunami features. Some relationships are easy to spot: tsunami amplitudes from earthquakes are proportional to the product of fault area and average slip ($A\Delta u$); tsunami amplitudes decrease with earthquake depth via the e^{-kd} terms. The ε_{ij} provide the dependence of tsunami amplitude and azimuthal radiation pattern on source type. Equation 10 says that tsunami from point sources radiate in azimuthal patterns no more intricate than $\sin 2\theta$ or $\cos 2\theta$.

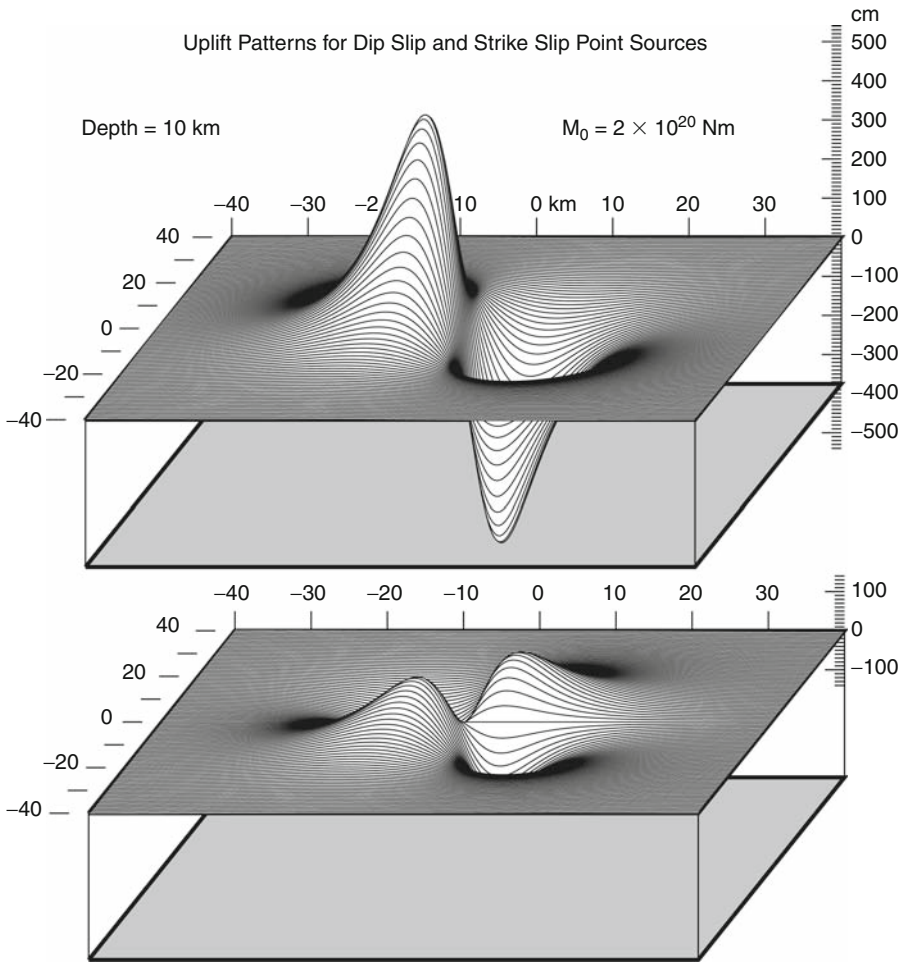
Figure 3 shows 5 h of tsunami waveforms calculated from (9) at distances of r = 200, 500, 1,000, 2,000 km from dip slip ($M_{yz} = M_{zy} = 1$) and strike slip ($M_{xy} = M_{yx} = 1$) point sources of magnitude $M_w = 7.5$ ($\Delta uA = 3.98 \times 10^6$ m^3. See Table 1) buried at 10 km depth. Sea waves from these sources have radiation patterns of $\sin\theta$ and $\sin 2\theta$, respectively. I compute the waveforms in Figure 3 at the azimuth of maximum strength, $\theta = 90°$ and $\theta = 45°$. Frequency dispersion, with the long periods arriving first, is the most conspicuous feature of the waveforms. Tsunami onset rapidly. They reach maximum height in the first few cycles and then decay slowly over an hour or more. Even for this large earthquake, tsunami beyond 500 km distance reach just a few centimeters – hardly killer waves. Note that if the observation direction was west versus east (Figure 3, *top*), or northwest versus northeast (Figure 3, *bottom*), the tsunami waveforms would be inverted. Whether tsunami onset in a withdrawal or an inundation at a particular shoreline strictly depends on the style of faulting and the relative positions of the shore and the fault.



Tsunami, Figure 3 Synthetic record sections of vertical tsunami motions at distances of 200, 500, 1,000, and 2,000 km from point dip slip (*top*) and strike slip (*bottom*) earthquakes of magnitude $M_w = 7.5$ and depth 10 km. Time runs for 5 h and the peak amplitude of each trace is given in centimeters at the right. The lower half of the focal sphere and azimuth of observation θ are shown toward the right. For other directions, the waveforms should be scaled by $\sin\theta$ and $\sin 2\theta$, respectively.

Figure 3 demonstrates that for point sources, dip slip earthquakes produce three or four times larger tsunami than strike slip earthquakes of equal moment. The differences in generation efficiency are understood most easily by considering directly the seafloor deformation patterns $u_z^{\text{bot}}(\mathbf{r}_0)$. I find these by setting $t = 0$ and $h = 0$ in (9) so $\cos\omega(k)t/\cosh(kh) = 1$. Figure 4 pictures the uplift patterns for the two faults of Figure 3 where two ($\sin\theta$) and four-lobed ($\sin 2\theta$) deformations spread over a region 40 km wide. The most striking contrast in the fields is maximum vertical displacement – 1.4 m for the strike slip versus 5.4 m for the dip slip. It is no coincidence that the ratio of maximum uplift for these two faults replicates the ratio of tsunami heights in Figure 3. After all, vertical seafloor deformation drives tsunami and vertical deformation is controlled largely by the rake of the slip vector $\hat{\mathbf{a}}$. Strike slip faults have a rake of 0° or 180°. Dip slip faults have rake equal ±90°. Further simulations show that, excepting very shallow, nearly horizontal faults, dip is not a terribly significant factor in tsunami production.

The sea surface cross sections in Figure 5 (*left* and *middle*) chronicle the birth and early life of a sea wave spawned by M7.5 thrust earthquakes on 45° dipping planes. In these figures, I replace the idealized point sources of Figure 3 with faults of typical dimension (L = 89 km, W = 25 km, and Δu = 1.78 m; see Table 1). In Figure 5 (*left*) the fault reaches to the sea floor. In Figure 5 (*middle*) the fault stops 30 km down. Soon after the earthquake, the sea surface forms "dimples" similar to those on the deformed sea floor. The sea surface dimples, however, are smoother and a bit lower in amplitude because of the $1/\cosh(kh)$ low-pass filtering effect of the ocean layer. After a time roughly equal to the dimension of the uplift divided by tsunami speed $\sqrt{gh}$, the leading edges of the wave organize and begin to propagate outward as expanding rings (Figure 5, *right*). Early on, the wave appears as a single pulse. Characteristic tsunami dispersion begins to be seen only after 10 or 20 min. Consequently, for shorelines close to tsunami sources, seismic sea waves arrive mostly as a single pulse. For distant shorelines, sea waves arrive with many oscillations, dispersion having spread out the initial pulse.



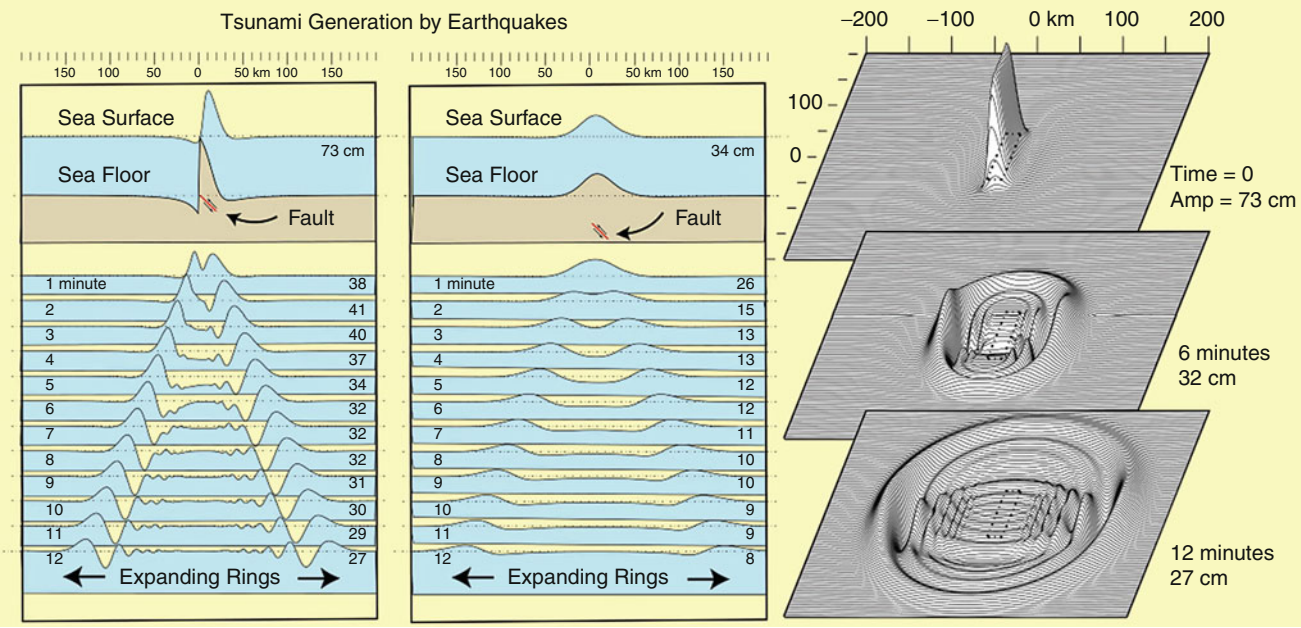
Tsunami, Figure 4 Static vertical displacements of the seafloor for the dip slip (*top*) and strike slip (*bottom*) earthquake point sources that generated the tsunami of Figure 3. Maximum excursions of the seafloor are 5.4 and 1.4 m, respectively.

The e^{-kd} terms in the tsunami excitation functions (10) let shallow earthquakes excite higher frequency tsunami than deep earthquakes (compare Figure 5 *left*, with Figure 5 *middle*). Higher frequency waves travel more slowly than longer period waves, so high frequency waves contribute to peak tsunami height only while it propagates as a single pulse. After a few hundred kilometers of travel, high frequency waves drift to the back of the wave train (Figure 5, *left*) and no longer add to the tsunami maximum. At 200 km distance, the shallow earthquake generates a wave about 3 times larger than the 30-km deep event. If you track the waves out to 2,000 km, however, you would find that the extra high frequencies in the shallow event will have fallen behind and that the maximum wave heights for the two events would be nearly equal. Beyond 2,000 km distance, any earthquake depth less than 30 km appears to be equally efficient in tsunami genesis.

Faults of finite size, like those in Figure 5, radiate tsunami in distributions more complex than the $\sin\theta$ and $\sin 2\theta$ patterns from point sources. The largest earthquakes have fault lengths of several hundred kilometers. Simulations show that long earthquake faults preferentially emit tsunami in a tight beam perpendicular to the fault strike, regardless of the focal mechanism. This preferential beaming simplifies tsunami forecasting because it tells us which direction to look for the biggest waves.

Tsunami excitation from submarine landslides

Earthquakes parent most tsunami, but other things do too. For instance, earthquake shaking often triggers landslides. If the slide happens under the sea, then tsunami may form. Consider a seafloor landslide confined in a rectangle of length L and width W. Let a constant uplift u_0 start along one width of the rectangle and run down its length ($\hat{\mathbf{x}}$ direction, say) at velocity v_r, that is, $\tau(\mathbf{r}_0) = x/v_r$. Placing this $u_z^{\text{bot}}(\mathbf{r}_0)$ and $\tau(\mathbf{r}_0)$ into (7a, 7b), I find the tsunami from this uplift source at observation point $\mathbf{r}$ and time t to be



Tsunami, Figure 5 *Left*. Cross sections of expanding tsunami rings from a M7.5 thrust earthquake. The fault strikes north–south (into the page) and the sections are taken east–west. Elapsed time in minutes and maximum amplitude in centimeters are given at the left and right sides. *Middle*. Cross sections of expanding tsunami rings from the same M7.5 thrust earthquake now buried 30 km. Deeper earthquakes make smaller and longer wavelength tsunami. *Right*. Views of the expanding tsunami rings from the shallow earthquake case at t = 6 and 12 min. The dashed rectangle in the center traces the surface projection of the fault. For large earthquakes, nearly all tsunami energy beams perpendicular to the strike of the fault (*toward the left and right in this picture*).

$$u_z^{\text{surf}}(\mathbf{r},t) = \frac{u_0 LW}{4\pi^2} \operatorname{Re}\int_k d\mathbf{k}\, \frac{e^{i(\mathbf{k}\cdot\mathbf{r}-\omega(k)t)}e^{-iX(\mathbf{k})}}{\cosh(kh)}\,\frac{\sin X(\mathbf{k})}{X(\mathbf{k})}\,\frac{\sin Y(\mathbf{k})}{Y(\mathbf{k})} \tag{12}$$

where

$$X(\mathbf{k}) = \frac{kL}{2}\left(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}} - c(k)/v_r\right) \quad \text{and} \quad \mathrm{Y}(\mathbf{k}) = \frac{\mathrm{kW}}{2}\left(\hat{\mathbf{k}}\cdot\hat{\mathbf{y}}\right).$$

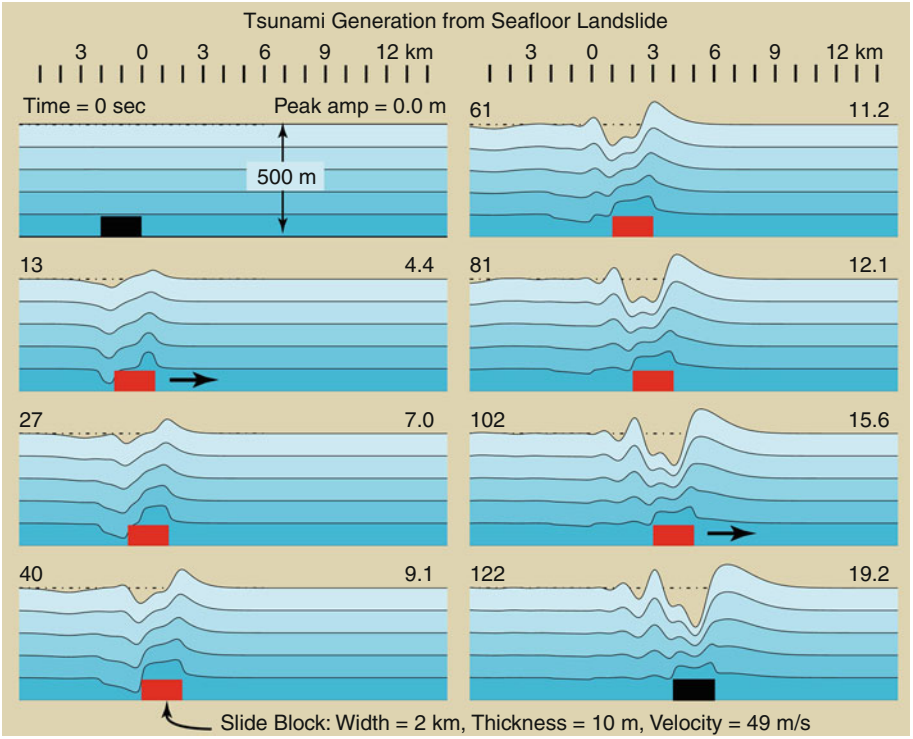
The $X(\mathbf{k})$ and $Y(\mathbf{k})$ factors, because they depend on the relative positions of the observation point and the landslide source, instill radiation patterns to the tsunami much like the ε_{ij} do for earthquakes.

By adding a trailing excavation source to the uplift source (12), I simulate tsunami from a sliding submarine block. Figure 6 pictures in cross section the waves stirred from a 2-km wide block, 10-m thick sliding at 49 m/s for 6 km along the bottom of an ocean 500 m deep. As the block moves, the water must go around the obstruction. Water gets pushed up in front and drawn down behind the slide block. Experiments reveal that, depending on the aspect ratio of the block and the ratio of slide velocity to the tsunami phase velocity, significant beaming and amplification of the tsunami are possible. In particular, when the slide velocity v_r approaches the tsunami speed $c(k) \approx \sqrt{gh}$, $X(\mathbf{k})$ is nearly zero for waves traveling in the slide direction. In this direction, the waves from different parts of the landslide arrive nearly "in phase" and constructively build. Figures 6 and 7 highlight the beaming and amplification effects. In both cases, witness the large tsunami pulse sent off in the direction of the slide. (Quicktime movies of Figures 6 and 7 and many other Figures in this paper are available for viewing; see Table 2.) Submarine landslides are prime suspects in the creation of "surprise tsunami" from small or distant quakes. Surprise landslide tsunami might initiate well outside of the earthquake uplift area or be far larger than expected given the magnitude of the quake. Historical examples of surprise landslide tsunami include the 1929 Grand Banks and the 1998 Papua New Guinea waves.

Tsunami excitation from impacts

In addition to tsunami from earthquakes and landslides, another class holds interest – those generated from surface detonations, explosive volcanoes, and asteroid strikes. I call waves from these sources "impact tsunami." Imagine that the initial stage of cratering by an impact excavates a radially symmetric, parabolic cavity of depth D_C, and radius R_C:

$$u_z^{\text{impact}}(\mathbf{r}) = D_C(1 - r^2/R_C^2) \qquad r \le \sqrt{2}R_C = R_D$$



Tsunami, Figure 6 Tsunami produced by a submarine landslide. The slide block is 10 m thick and 2 km wide. The slide starts to the left and runs to the right at 49 m/s. Note the amplification of the wave in the slide direction.

$$u_z^{\text{impact}}(\mathbf{r}) = 0 \qquad r > \sqrt{2}R_C = R_D$$

Based on Equation 8, the impact tsunami at observation point **r** and time t is

$$u_z^{\text{surf}}(\mathbf{r}, t) = \int_0^\infty \frac{k\,dk}{2\pi} \cos[\omega(k)t] J_0(kr) F_0(k), \qquad (13)$$

where

$$F_0(k) = \int_{\mathbf{r}_0} r\,d\mathbf{r}_0\, u_z^{\text{impact}}(\mathbf{r}_0) J_0(kr_0) = \frac{2\pi D_C R_D}{k} J_3(kR_D).$$

The principal distinction between (8) and (13) is the absence in the latter of the $1/\cosh(kh)$ low-pass ocean filter. Impacts crater the surface of the ocean not the sea floor, so this filter does not come into play.

If we restrict ourselves to asteroid-induced tsunami and suppose that the depth of the impact cavities equals 1/3 their diameter $d_c = 2R_C$, then D_c relates scales with the density, velocity, and radius of the impacting asteroid as

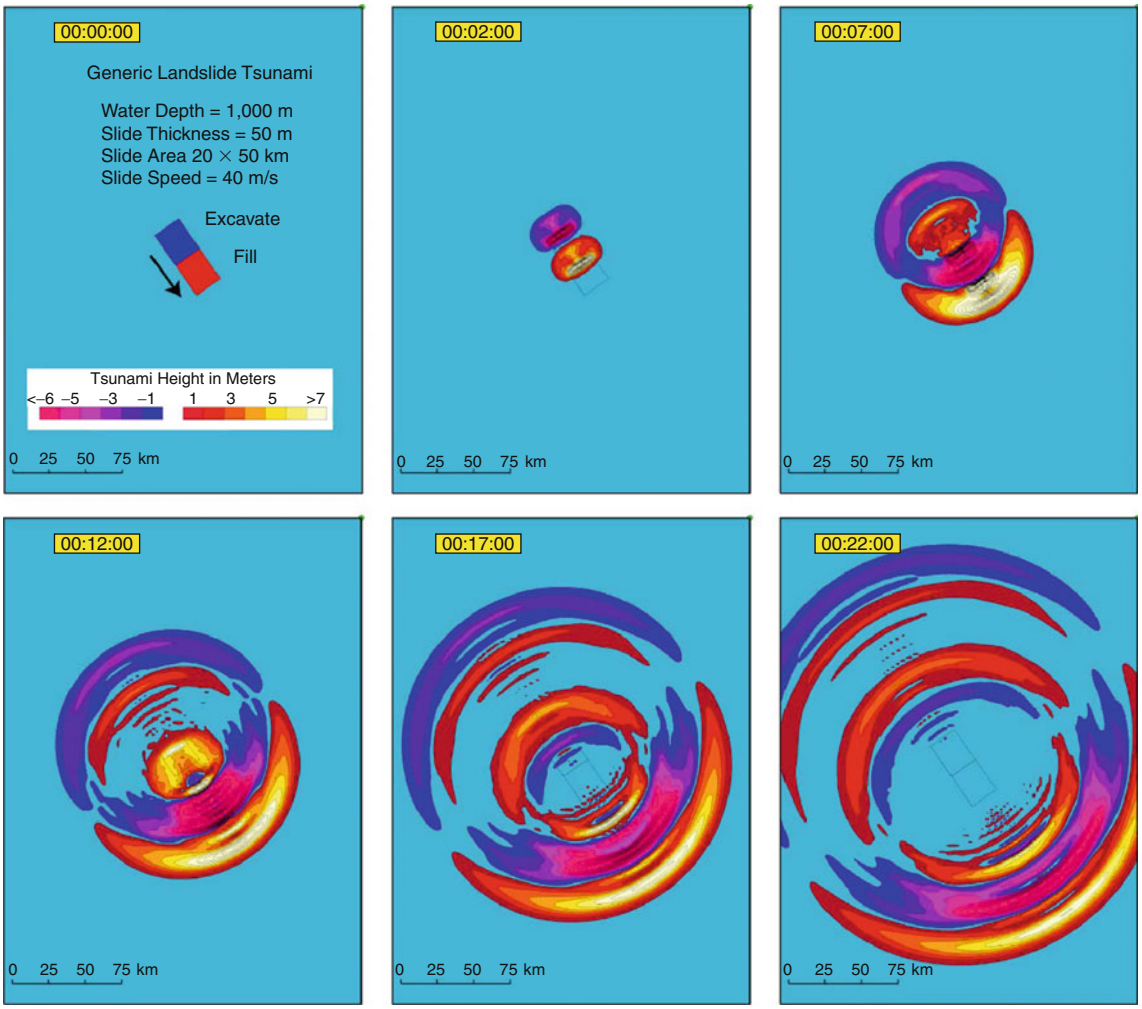
$$D_C = d_c/3 = (8\varepsilon\rho_I V^2{}_I/9\rho_w g)^{1/4} R_I^{3/4} \qquad (14)$$

In (14), ε is the fraction of the kinetic energy of the asteroid that goes into the tsunami wave. High velocity impact experiments suggest that $\varepsilon \approx 0.15$. With $\rho_I = 3$ gm/cm^3, $\rho_w = 1$ gm/cm^3, and $V_I = 20$ km/s, (14) returns crater depths of 1,195, 2,010, and 4,000 m for asteroids of radius $R_I = 50$, 100, and 250 m.

Figure 8 plots cross sections of the expanding rings of tsunami waves induced by the impact of a 200-m diameter asteroid at 20 km/s as computed by Equation 13. Within 100 km of ground zero, tsunami from moderate size (100–250 m) asteroids have heights of many hundreds of meters and dwarf the largest (10–15 m) waves parented by earthquakes. Fortunately, impact tsunami fall on the steep-sloped part of the group velocity curve (Figure 1) so they decay faster with distance than earthquake tsunami (more like r^{-1} versus $r^{-3/4}$ for earthquake tsunami). Figure 9 illustrates how dispersion quickly pulls apart the initial impulse into dozens of waves. At 1,000 and 3,000 km, the tsunami in Figure 8 would decay to 6 and 2 m amplitude, respectively – still a concern, but not catastrophic. For perspective, asteroids with diameters >200 m impact Earth about every 10,000 years, far less frequently than great M9 earthquakes that strike the planet once in 25 years or so.

Tsunami propagation

In uniform depth oceans, tsunami propagate out from their source in circular rings (e.g., Figure 9) with ray paths that look like spokes on a wheel. In real oceans, tsunami



Tsunami, Figure 7 Map view of a tsunami from a landslide block similar to Figure 6. Note the dipole form of the wave in its early stages and the beaming in the southeast direction of the slide.

speeds vary place to place (even at a fixed frequency) so tsunami ray paths refract and become bent. Consequently, in real oceans, both tsunami travel time and amplitude have to be adjusted relative to their values in uniform depth ones. To propagate tsunami in real oceans, I find it best to keep tsunami "mode-like" versus depth but "ray-like" horizontally. I first transform the various integrals over wavenumber (7, 8, 9, etc.) to integrals over frequency because wave frequency, not wave number, is conserved throughout. Using the relations $u(\omega) = d\omega/dk$ and $c(\omega) = \omega/k(\omega)$, I find that tsunami vertical motions from (13) for instance, are to a good approximation:

$$u_z^{\text{surf}}(\mathbf{r},t) = \int_0^\infty \frac{\cos(\omega t)\mathrm{d}\omega}{2\pi k(\omega)u(\omega)} J_0(\omega T(\mathbf{r},\omega))F_0(k(\omega)) G(\mathbf{r})S_{\mathrm{L}}(\omega,\mathbf{r}). \quad (15)$$

In (15) the travel time of waves of frequency ω has been changed from $r/c(\omega)$ to

$$T(\mathbf{r},\omega) = \int_{\text{raypath}} \mathrm{d}r/c(\omega,h(\mathbf{r})), \quad (16)$$

where the integral path traces the tsunami ray from the source to the observation point. Equation 15 also incorporates a new shoaling factor $S_{\mathrm{L}}(\omega,\mathbf{r})$ that accounts for wave height changes due to water depth and new ray geometrical spreading factor $G(\mathbf{r}) \leq 1$ that takes into consideration the reduction of wave amplitudes into shadow zones.

Real geometrical rays for tsunami can be very messy (*top left*, Figure 10). Rather, I employ "network rays" that are simply minimum time paths between source and receiver (*top right*, Figure 10). Unlike geometrical rays, network rays possess neither caustics nor shadow zones.

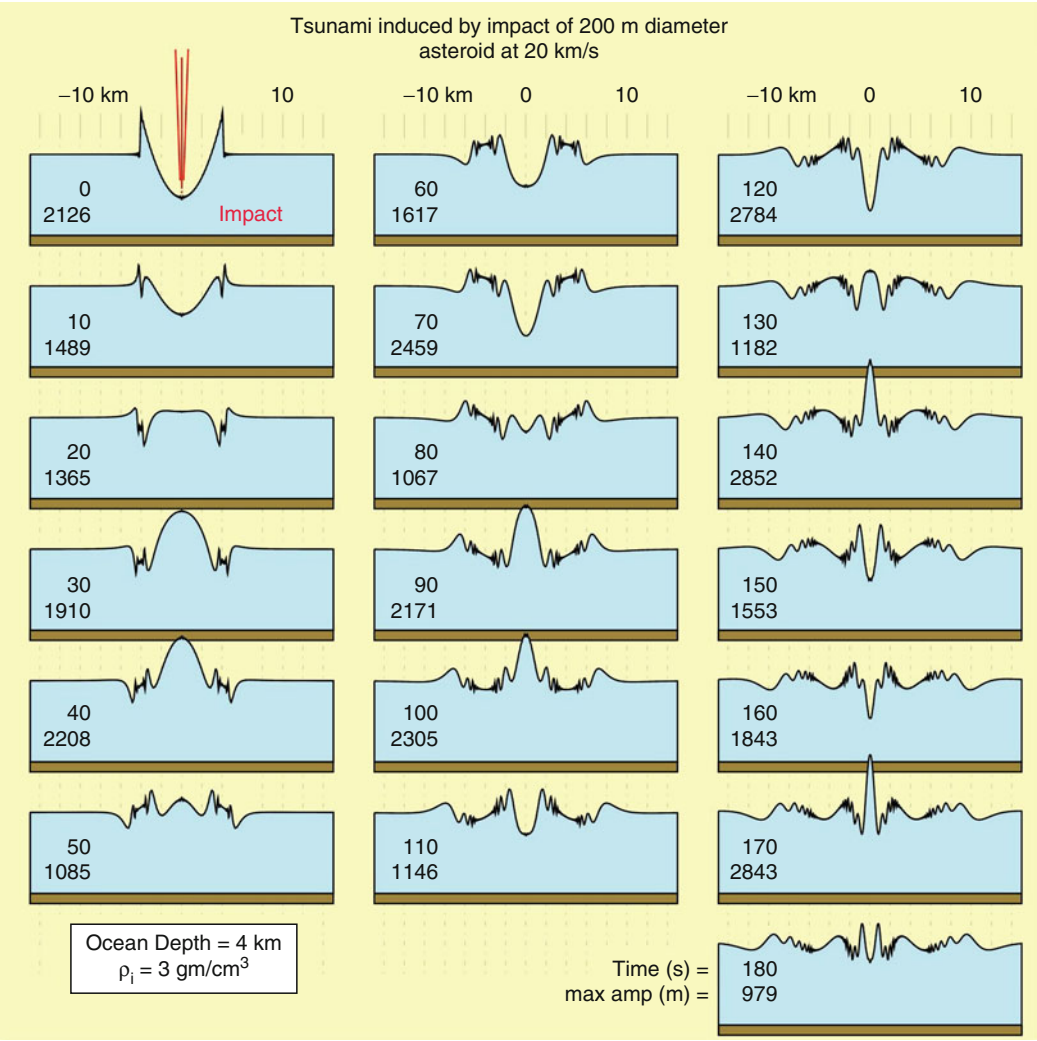
Tsunami, Table 2 Quicktime movie links to many of the figures in this article, plus a few others. If you want your own copy: Play the Movie in a Browser, then click "Save to Disk."

Figure 6	http://es.ucsc.edu/~ward/landslide(1_big).mov
Figure 7	http://es.ucsc.edu/~ward/generic-slide.mov
Figure 8	http://es.ucsc.edu/~ward/imp-parab1.mov
Figure 9	http://es.ucsc.edu/~ward/impact-example-3d.mov
Figure 13	http://es.ucsc.edu/~ward/RU(1200s-20m-0.5).mov
Figure 14	http://es.ucsc.edu/~ward/galveston-2m-close.mov
	http://es.ucsc.edu/~ward/galveston-2m.mov
Figure 15	http://es.ucsc.edu/~ward/indo-3D3.mov
	http://es.ucsc.edu/~ward/indo-3D2.mov
Figure 16	http://es.ucsc.edu/~ward/LP-3D.mov
Figure 17	http://es.ucsc.edu/~ward/guinea(270).mov
	http://es.ucsc.edu/~ward/ire-nn.mov
Figure 18	http://es.ucsc.edu/~ward/samoa-92909(c).mov

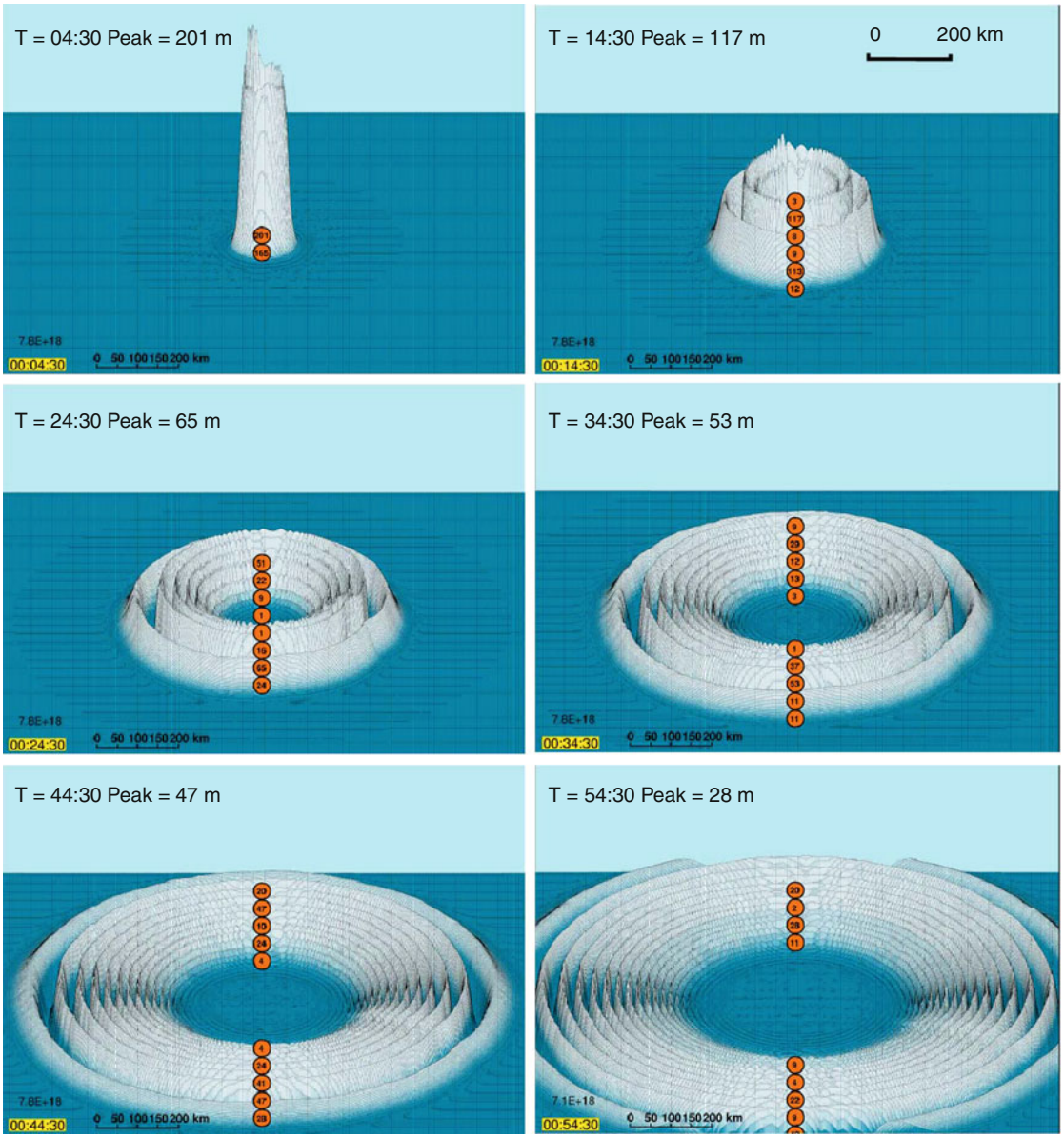
To compute the ray geometrical spreading factor $G(\mathbf{r})$, I replace the curvature of the wave front with curvature of the ray in the path integral (Figure 10, *bottom row*):

$$G(\mathbf{r}) = \exp\left[-\frac{1}{2}\int_{\text{raypath}} \frac{\partial\alpha(s)}{\mathrm{d}w}\mathrm{d}s\right] \Rightarrow \exp\left[-\frac{1}{2}\int_{\text{raypath}} \left[\frac{\partial\theta(s)}{\mathrm{d}s}\right]\mathrm{d}s\right]. \tag{17}$$

In this approach the distinction between wave refraction and diffraction blurs.



Tsunami, Figure 8 Computed tsunami induced by the impact of a 200-m diameter asteroid at 20 km/s. The waveforms (shown at 10 s intervals) trace the surface of the ocean over a 30 km cross section that cuts rings of tsunami waves expanding from the impact site at $x = 0$. Maximum amplitude in meters is listed to the left.



Tsunami, Figure 9 Expanding rings of an impact tsunami in a 3D-like style. Note the strong wave spreading due to dispersion.

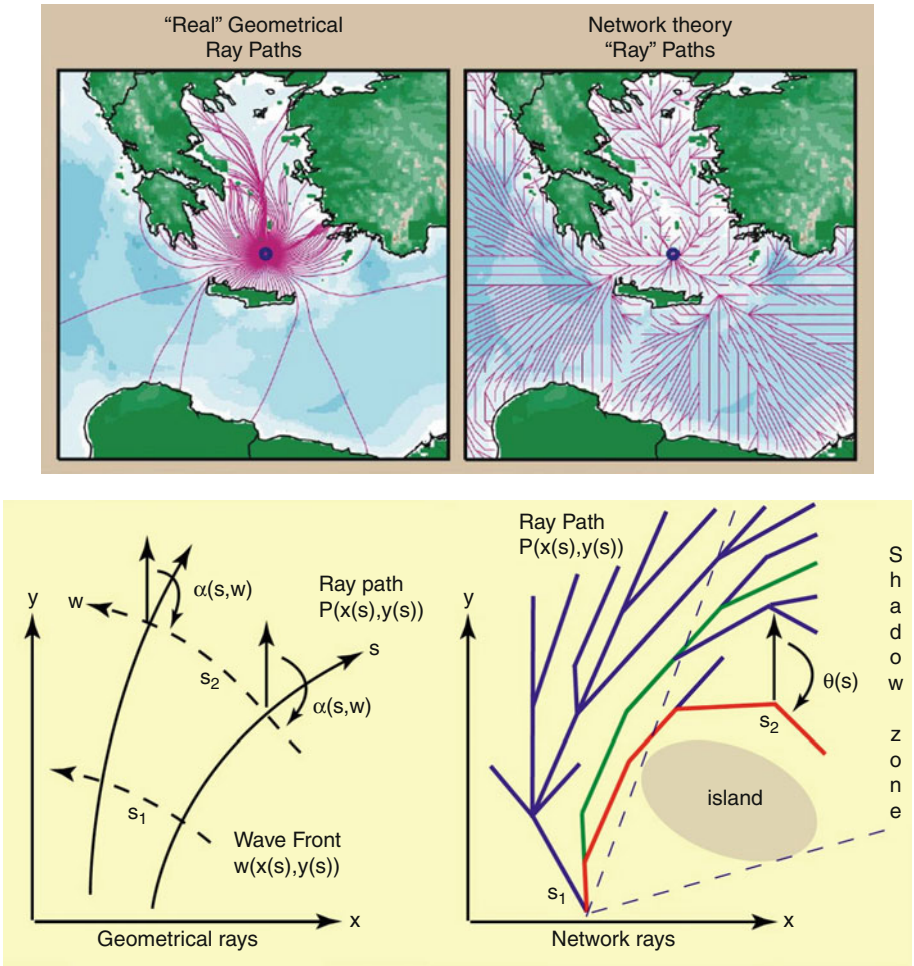
Tsunami shoaling and run-up

Shoaling. Toward shore, real oceans shallow and the waves carried on them amplify in a process called *shoaling*. Shoaling is easy to understand. The tsunami velocity depends on the ocean depth so as water shallows, tsunami waves slow down. Because their frequency is fixed, the wavelength of a slowing tsunami decreases (Figure 11, *inset*) and its energy compresses horizontally. Secondly, because a tsunami occupies the entire water column, as it enters shallow water its energy also gets compressed vertically. The only way for the compressing wave to maintain the same energy flux is for it to grow in amplitude. For the shoaling factor in (15), linear theory gives:

$$S_{\mathrm{L}}(\omega, \mathbf{r}) = \sqrt{\frac{u(\omega,h_0)}{u(\omega,h(\mathbf{r}))}}. \tag{18}$$

Shoaling amplification depends on the ratio of group velocity at the nucleation-site and the coast-site (ocean depths h_0 and $h(\mathbf{r})$, respectively). As does $G(\mathbf{r})$, S_{L} reverts to one in oceans of uniform depth.

Figure 11 pictures a shoaling tsunami wave of 150 s period. Initially, a unit height wave comes ashore from

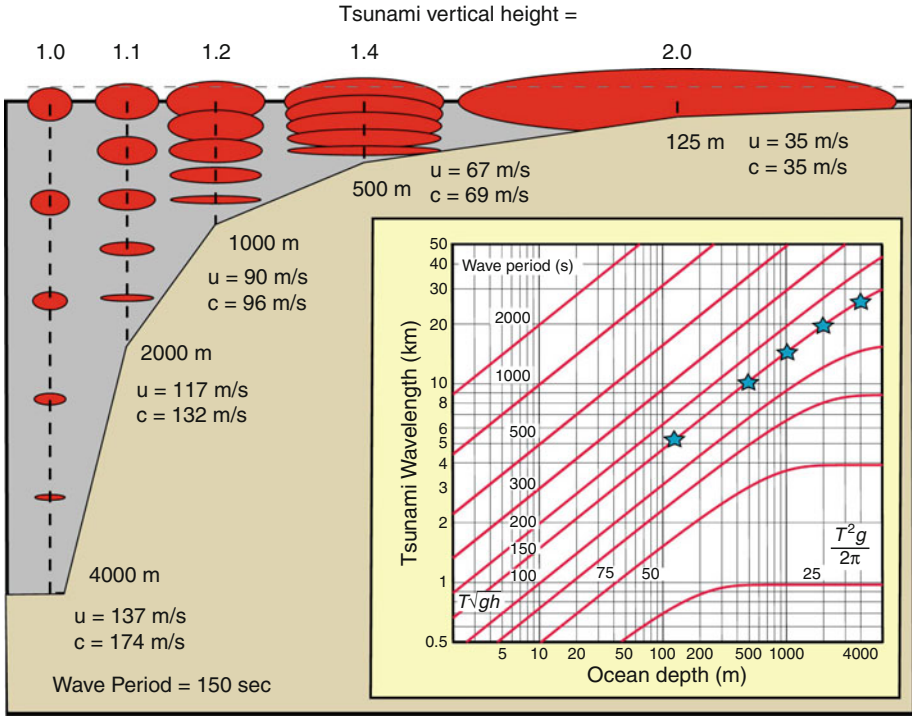


Tsunami, Figure 10 Real tsunami rays (*top left*) versus network rays (*top, right*). I track the curvature of the ray (*right, bottom*) rather than the curvature of the wave front (*left, bottom*) to account for wave size losses as tsunami propagate around obstacles.

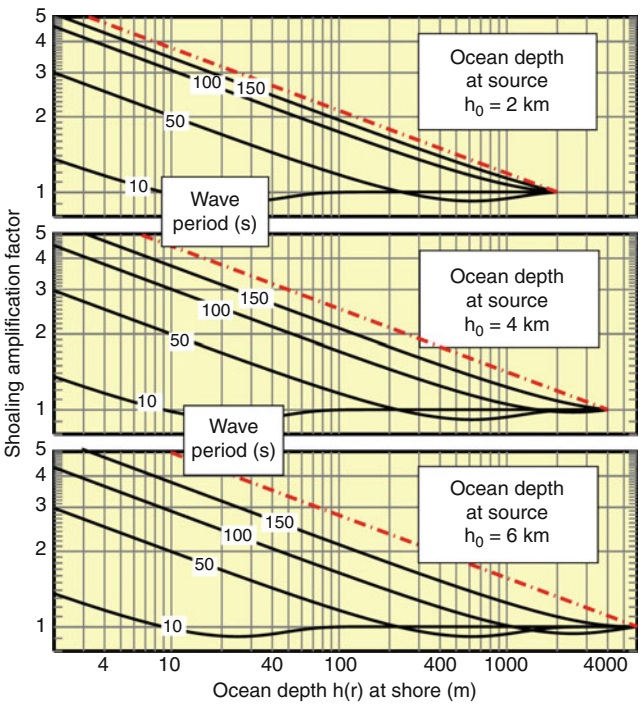
4,000 m of water at the left. As the water shallows, the velocity of the wave decreases and the wave grows in amplitude. By the time it reaches 125 m depth it has slowed from 137 m/s group velocity to 35 m/s and grown in height by a factor of 2. Figure 12 plots (18) as a function of coast-site depth for sea-wave periods from 10 s, and ocean depths of 2, 4, and 6 km. Because their energy does not occupy the entire water column, beach waves at 10 s period do not amplify much (perhaps 50%) in shoaling. Tsunami waves (100–2,000 s period) experience much stronger shoaling amplification – about 3–6 over a wide range of conditions. For waves of period greater than 250 s, $u(\omega,h_0) = (gh_0)^{1/2}$, $u(\omega,h(r)) = (gh(\mathrm{i}))^{1/2}$, and the shoaling factor reduces to Green's Law, $S_L = (h_0/h(r))^{1/4}$ (dashed red line in Figure 12).

Run-Up. Run-up is the final phase of tsunami life. The run-up phase begins when the approaching tsunami shoals to an amplitude roughly equal to the water depth and the wave begins to break. Run-up also covers the inundation phase where the water runs over land and reaches its maximum excursion above sea level. Run-up is the most complicated phase of tsunami life; nevertheless, "Will I get wet?" is the question everyone wants answered, so tsunami scientists have to take their best shot at carrying the waves to the last mile.

One approach to run-up is wave tank or computer experiments like Figures 13 and 14, where waves of various sizes and periods are sent onto model beaches. These experiments reveal a wide range of behaviors depending on wave period, wave size, wave direction, number of waves, beach slope, and beach friction. In real-world situations, this information is not known well, so the direct applications of the experimental results are limited. I follow an alternative approach that takes a broad brush view of run-up, but allows for considerable random uncertainty in the outcome. Using the methods above, I propagate tsunami to a position close to shore in shallow water where it can be considered a long wave, but not so



Tsunami, Figure 11 Effect of shoaling on tsunami eigenfunctions. The shallowing ocean near shore concentrates wave energy into smaller and smaller volumes and tsunami grow in response.



Tsunami, Figure 12 Shoaling amplification factor for ocean waves of various frequencies and source depths. The *dashed red line* is Green's Law.

shallow that its amplitude $A(h)$ exceeds water depth h there. I estimate run-up height η, on shore by:

$$\eta = A(h)^{4/5} h^{1/5} \tag{19}$$

Because $A(h) < h$, run-up always exceeds the offshore wave amplitude. Considering the large uncertainties involved, I view run-up at any field site to be a random variable drawn from a distribution function with (19) being its mean and its spread ranging from 1/2 to 2 times the mean. It does not matter at which depth h is used to evaluate (19), provided the water is shallow enough for the tsunami to act like a long wave. According to Green's Law, a long wave with amplitude $A(h_1)$ in water of depth h_1, has an amplitude $A(h_2) = A(h_1)\,(h_1/h_2)^{1/4}$ in water of depth h_2. Run-up estimate at η_2 at depth h_2

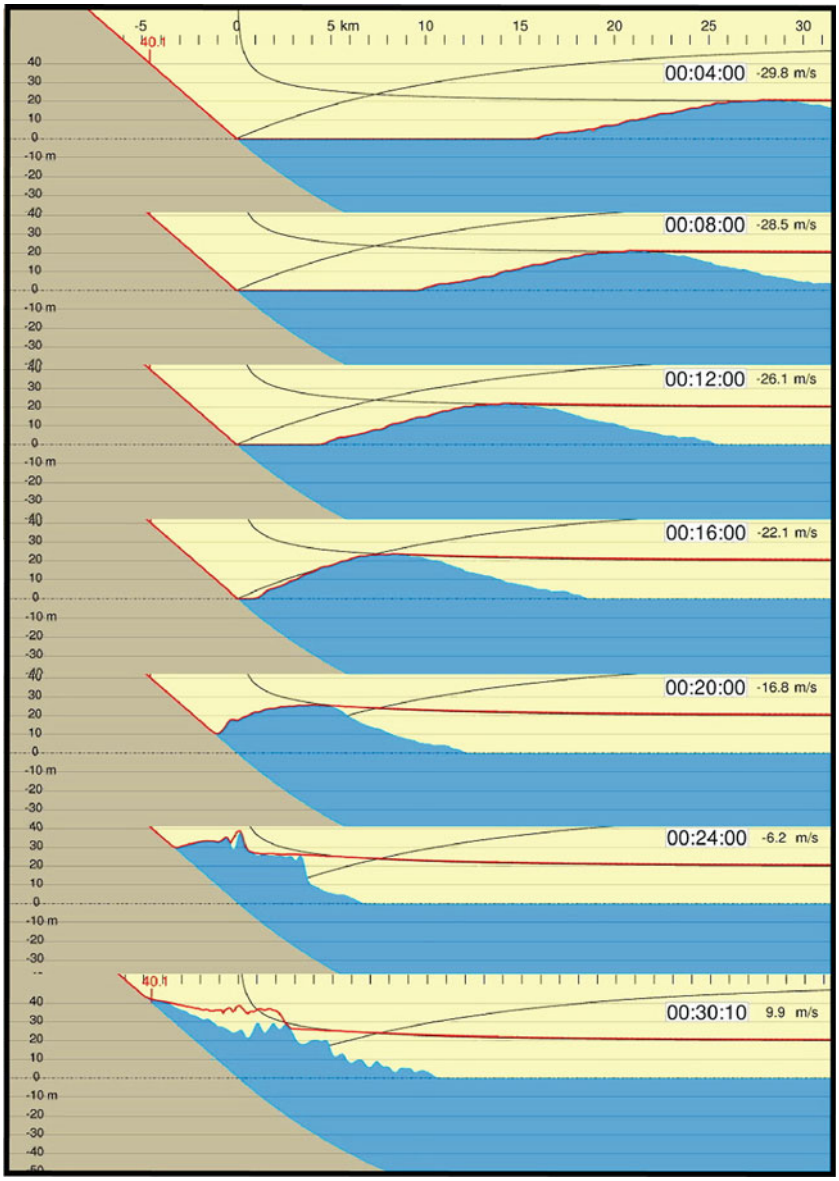
$$\eta_2 = A(h_2)^{4/5} h_2^{1/5} = \left[A(h_1) \left[\frac{h_1}{h_2} \right]^{1/4} \right]^{4/5} h_2^{1/5}$$

$$= A^{4/5}(h_1) {h_1}^{1/5} = \eta_1$$

equals estimate η_1 at depth h_1.

Tsunami samples

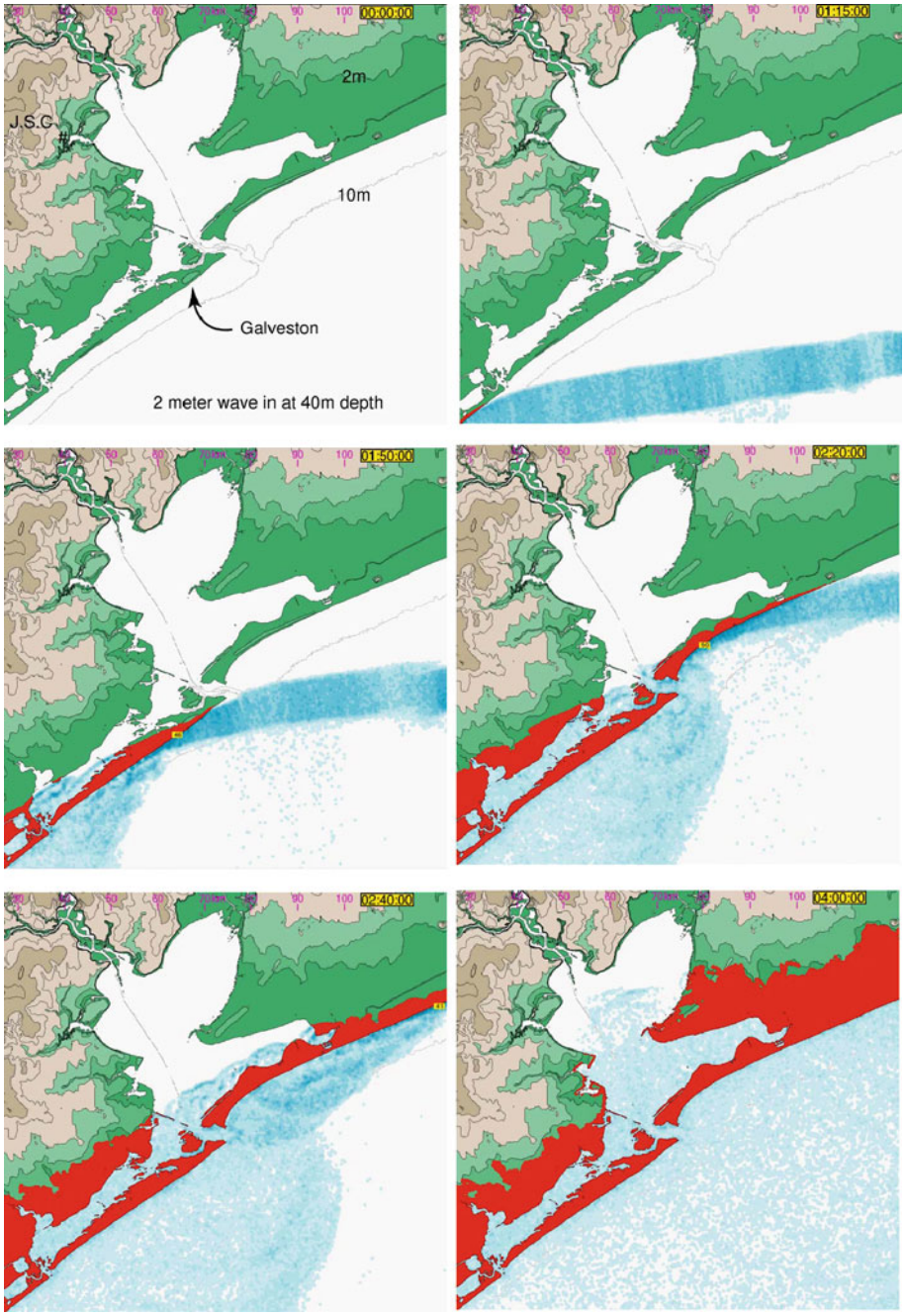
Sumatra earthquake 12/26/2004. Figure 15 snapshots a simulation of the 2004 Sumatra tsunami. As modeled



Tsunami, Figure 13 Run-up and inundation of a 20 m high, 1,200 s duration wave on a steep 0.5° slope. On steep beaches breaking has little time to operate so run-up amplifications are large.

by several rectangular patches, the megathrust fault that sourced the wave trends north–south for several 100 km (Figure 15 *top left*). As the example in Figure 5 predicted, nearly all of the wave energy in the 2004 tsunami emits perpendicular to the fault – Sumatra and Thailand toward the east, and India and Sri Lanka toward the west. Not surprisingly, these locations suffered the lion's share of tsunami damage. I believe that the very large run-up near Banda Aceh was due to concentrated slip on a thrust ramp splaying up from the megathrust just offshore at that latitude.

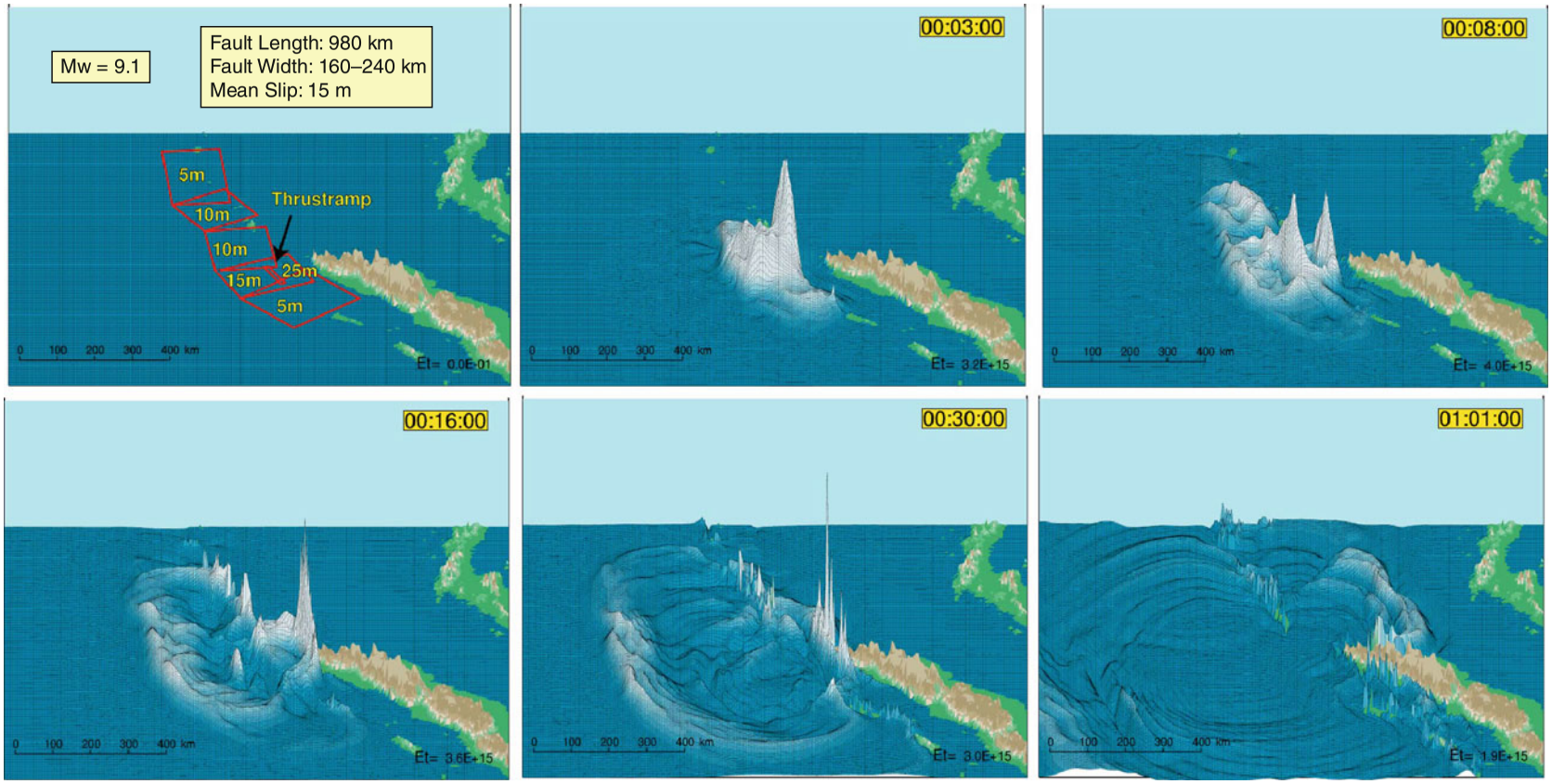
La Palma landslide tsunami. Like its fellows in the Canary Island chain, the volcanic Island of La Palma has experienced cycles of growth and partial collapse. Its flank last failed 550,000 years ago. Since then, new Cumbre Vieja volcano has filled the vacant scar. For the last 125,000 years this volcano grew steadily but in the last 7,000, its structure changed. It now mainly erupts along a north–south trending rift that splits the mountain in half. Moreover, during an eruption in 1949 a fault broke surface along the crest of the volcano and part of its western side slid 5 m down and toward the ocean. The volcano again may be showing initial stages of instability. While certainly collapse is not imminent and it may take many eruptive cycles over the next few thousand years to give it that final shove, the volcano will collapse so it is sensible to consider the consequences (Figure 16). If 500 km^3 of its



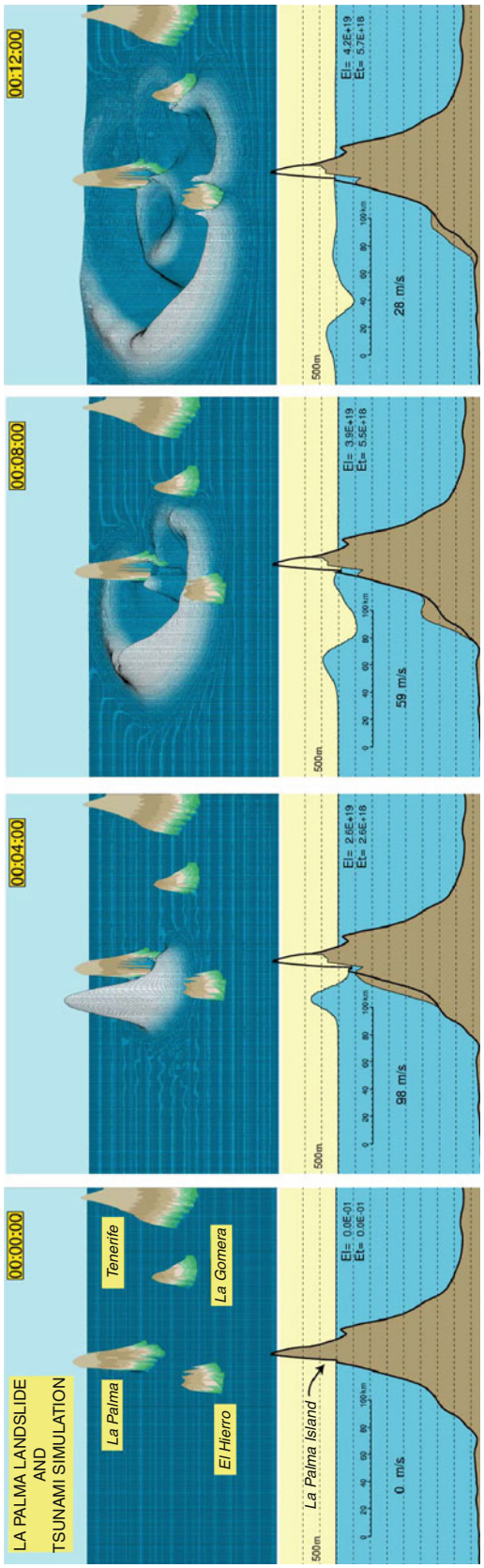
Tsunami, Figure 14 Simulation of 2 m tsunami wave beaching at Galveston Texas. Red area blankets the predicted inundation zone. Note the multiple wave reflections and wave interference that contribute to run-up complexity and randomness.

flank falls into the sea like it did last time, a La Palma collapse will spawn a mega tsunami hundreds of meters high locally that will buffet all coasts of the Atlantic with 10–30 m waves.

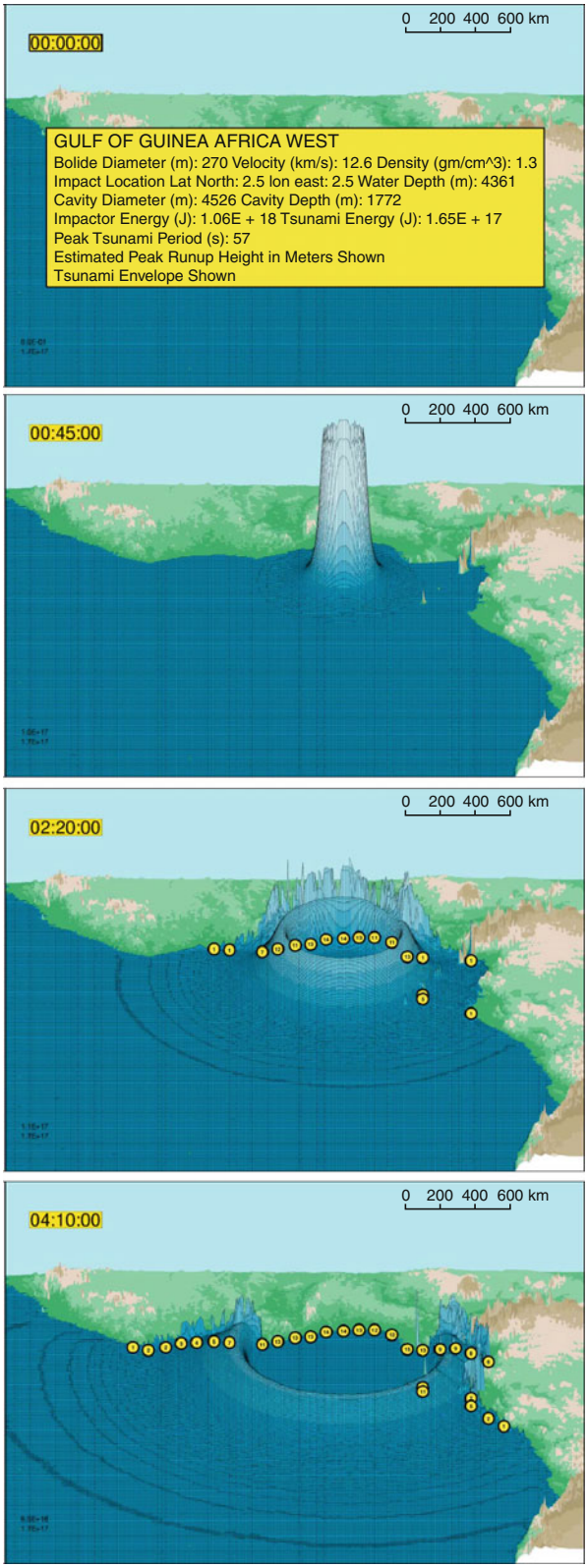
Gulf of Guinea impact tsunami. In a 2008 training exercise, the US Air Force considered the consequences of, and response to, an impact of a small comet into the Atlantic Ocean off the coast of Guinea. In this scenario, pictured in Figure 17, tsunami run-ups exceed 10 m along a 1,200 km stretch of coastline. This wave would knock out coastal shipping and oil production facilities for many months and trigger major ripple effects in the global economy. Unlike those Hollywood movies, currently there is no defense against an asteroid impact.



Tsunami, Figure 15 Simulation of the Sumatra Earthquake Tsunami of December 2004. Note the large wave directed toward the northern tip of Sumatra due to the small thrust ramp offshore there (*upper left*).



Tsunami, Figure 16 (*Top row*) First 12 min of the La Palma Tsunami. Initial wave heights exceed 500 m. (*Bottom row*) Landslide and tsunami in cross section. Just like the simple example in Figure 6, water gets pushed up in front of the slide and drawn down behind.



Tsunami, Figure 17 Guinea Impact Tsunami. Four-hour sequence of waves from the impact of a 270 m diameter icy bolide off the west coast of Africa.

Tsunami forecasting

Official approach: Perhaps the ultimate goal of tsunami research is forecasting. A forecast predicts the size of a particular tsunami given the knowledge that a potentially dangerous earthquake has occurred already. Tsunami travel at jet speeds, but for many places there may be several hours between the earthquake and the arrival of the sea waves. This time can be spent analyzing seismograms, estimating earthquake parameters, and forecasting the expected height of the oncoming wave with the aid of computer models and initial field reports. This is precisely the mission of the various warning centers like NOAA's Pacific Tsunami Warning Center (PTWC). The past decade has witnessed considerable progress in tsunami forecasting coupled with faster and more geographically selective warnings. By and large, the current tsunami prediction system functions well. Figure 18 pictures a nearly real-time prediction of tsunami run-up from the M8.1 Samoa Earthquake of September 29, 2009. Based on rapid calculations like these, the PTWC issued accurate advanced warnings and watches for this tsunami – at least for locations greater than 2,000 km distance from the quake where sufficient delay time existed.

The biggest uncertainty in the current procedures lay in fixing earthquake parameters quickly. For instance, in the first hours after a quake, it is not uncommon for estimates of earthquake magnitude to vary 0.3 or 0.4 units as more extensive seismic data arrive. Because earthquake magnitude or earthquake moment is the prime driver for tsunami (Equation 9), a 0.4 unit uncertainly makes for a factor of 2 or 3 uncertainly in tsunami height all else aside. Partly on this concern, the DART buoy network that provides real-time measurements of passing tsunami was expanded from 6 to 24 sites after the Sumatra earthquake. Information from the DART sensors provides a real-time "reality check" on computer simulations run rapidly on uncertain earthquake parameters.

Do-it-yourself approach: Apart from official predictions, readers of this article have enough information to produce a Do-It-Yourself tsunami forecast. The three step recipe needs: earthquake magnitude M, the ocean depth at source H_o, distance to the quake R, and mean ocean depth $\overline{H}$ between the earthquake and you.

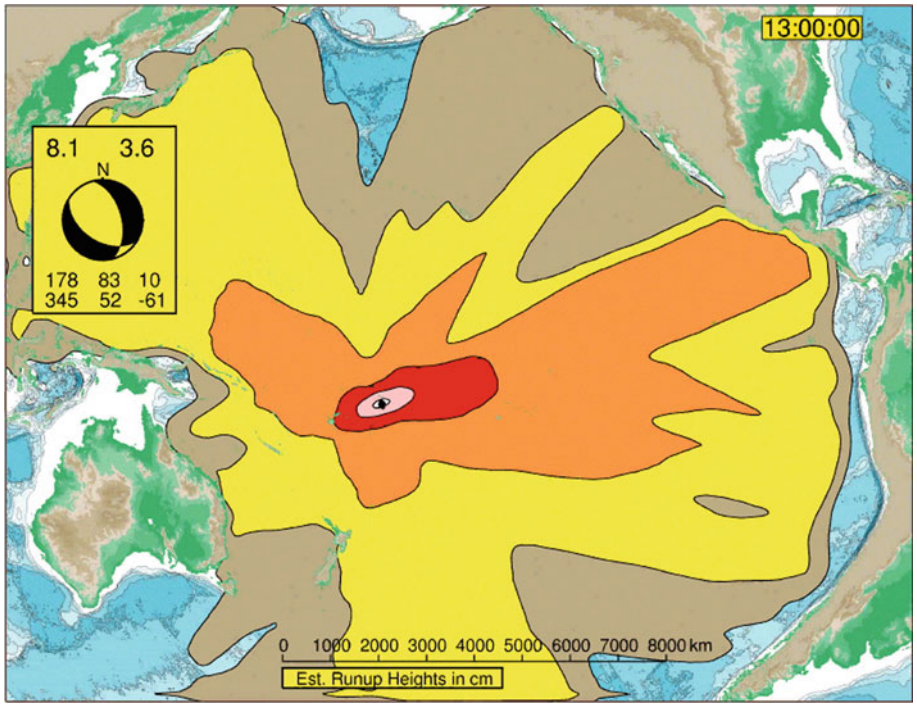
Step 1 – Find Initial Tsunami Amplitude A_0. Let us approximate initial tsunami height by:

$$A_0 = \frac{\alpha \Delta u(M)}{\cosh\left[\frac{4\pi H_0}{W(M)+L(M)}\right]}, \tag{20}$$

where α is the fraction of earthquake slip that transforms into tsunami-making uplift:

$$\alpha = (1 - \theta/180)\sin\theta|\sin\phi|,$$

with θ and ϕ being the fault dip and rake angles in degrees. Fault slip $\Delta u(M)$, width $W(M)$, and length $L(M)$ are functions of magnitude and can be read from Table 1 or computed from the formulas in its caption.



Tsunami, Figure 18 Nearly real-time prediction of run-up height from the 9/29/09 M8.1 Samoa Earthquake. Colors correspond to run-up heights: *Brown* 10–25 cm; *Yellow* 25–50 cm; *Orange* 50–100 cm; *Red* 1–2 m; *Pink* >2 m.

You recognize the 1/cosh as the low-pass filter due to the ocean layer. Unlike Equation 9, Equation 20 makes no assumption about radiation pattern, so in this sense it represents the worst case. The most efficient mechanism (surface breaking, pure dip slip on planes dipping near 45°) for tsunami generation gives $\alpha = 0.55$.

Step 2 – Correct for Propagation Loss *P*.

Tsunami waves decay as they travel due to geometrical spreading and frequency dispersion. For an ocean of constant depth $\overline{H}$, the propagation loss is roughly

$$P = \left(1 + \frac{2R}{L(M)}\right)^{-\psi}, \tag{21}$$

where

$$\psi = 0.5 + 0.575 \exp\left(-0.0175 \frac{L(M)}{\overline{H}}\right). \tag{22}$$

The first term in exponent (22) accounts for square root of distance losses due to geometrical spreading. The second term in (22) accounts for additional losses due to frequency dispersion with larger dimensioned tsunami decaying more slowly. Peak deep-water tsunami amplitude $A_d(R)$ at distance R is

$$A_d(R) = A_0 P = \frac{\alpha \Delta u(M)}{\cosh[4\pi H_0/(W(M) + L(M))]} \left(1 + \frac{2R}{L(M)}\right)^{-\psi} \tag{23}$$

Propagation loss (23) supposes an unobstructed wave path to the observation location. If the waves pass around headlands, squeeze through narrows, or cut into shadow zones, then (23) again overestimates tsunami size.

Step 3 – Correct for Shoaling and Run-Up.

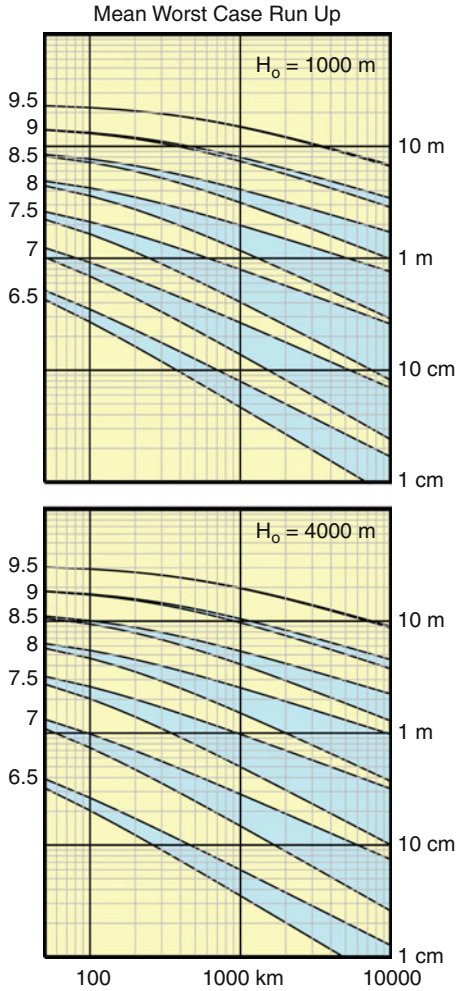
For long waves, the shoaling correction S_L (18) reduces to Green's Law $S_L = (H_0/H)^{1/4}$ and the shoaled tsunami amplitude $A_S(R)$ at distance R is

$$A_S(R) = A_0 P S_L = \frac{\alpha \Delta u(M)}{\cosh\left[\frac{4\pi H_0}{W(M)+L(M)}\right]} \left(1 + \frac{2R}{L(M)}\right)^{-\psi} \left[\frac{H_0}{H}\right]^{1/4}. \tag{24}$$

Applying (19) on (24) returns peak wave run-up height $\eta(R)$

$$\eta(R) = [A_0 P S_L]^{4/5} H^{1/5} = \left[\frac{\alpha \Delta u(M)}{\cosh\left[\frac{4\pi H_0}{W(M)+L(M)}\right]} \left(1 + \frac{2R}{L(M)}\right)^{-\psi}\right]^{4/5} H_0^{1/5}. \tag{25}$$

Figure 19 plots Do-It-Yourself prediction (25) for earthquake magnitudes 6.5–9.5, at distances R to 10,000 km, for $\overline{H}$ between 1,000 and 4,000 m, and $H_0 = 1{,}000$ m (*top*



Tsunami, Figure 19 Mean worst-case tsunami run-up height versus distance from earthquakes of magnitude 6.5–9.5. The blue areas include an allowance for mean ocean depth between 1,000 (*upper limit*) and 4,000 m (*lower limit*). Top and Bottom panels assume 1,000 and 4,000 m water depth at the source.

panel) and H_0 = 4,000 m (*bottom panel*). Because several steps above selected extreme assumptions, these curves represent a mean worst case. The worst worst case might be twice as large still. For $M = 8$, $H_0 = \overline{H} = 4{,}000$ m, Figure 19 (*bottom*) predicts run-ups of 3 m at 300 km distance, 2 m at 600 km distance, 1 m at 2,000 km distance, and 0.5 m at 5,000 km distance. These numbers compare well with the more rigorous calculations in Figure 18.

Summary

Tsunami are gravity-driven water waves caused mostly by undersea earthquakes, but sometimes they are sourced by submarine landslides, surface detonations, explosive volcanoes, and asteroid strikes. Tsunami differ from ordinary beach waves in that they have longer period, larger wavelength, and higher velocity. Because of shoaling effects, these characteristics make tsunami especially damaging because they grow in height proportionally more in approaching shore than do ordinary waves of equal size. This article provides mathematical expressions for tsunami excitation, propagation, and simplified steps to compute wave run-up height from tsunamigenic earthquakes.

Cross-references

Earthquakes and Crustal Deformation
Earthquakes, Source Theory
Seismic, Ray Theory
Surface Waves
Tsunami Watch and Warning Centers
Tsunami: Bay of Bengal

TSUNAMI: BAY OF BENGAL

Vineet Gahalaut
National Geophysical Research Institute (CSIR), Hyderabad, India

Definition

Bay of Bengal: Northeastern part of the Indian Ocean, which is about 2,000 km long and about 1,500 km wide, and is bordered on the west by India and Sri Lanka, on the north by Bangladesh, and on the east by Myanmar and Thailand; the Andaman and Nicobar Islands separate it from the Andaman Sea.
Tsunami: Gravity-driven water waves in the sea caused by submarine earthquakes, landslides, volcanic eruption, and impact meteorites.

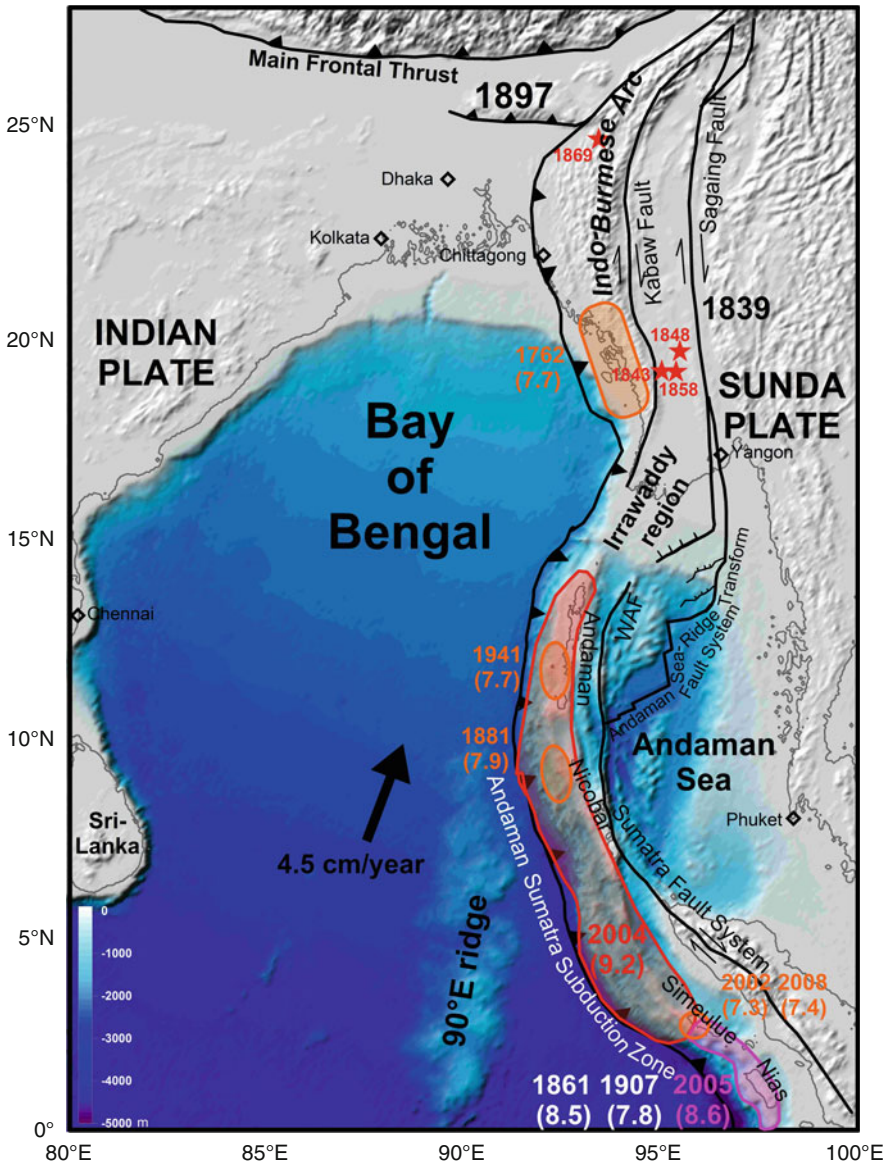
Introduction

The coastal region around the Bay of Bengal is one of the most populated regions of the world with population density of more than 1,000 people/km^2 in Bangladesh and region around it. Natural hazards in this region, namely, floods and cyclones, have turned into disasters making this region very vulnerable. The occurrence of the December 26, 2004, Sumatra-Andaman earthquake (Mw 9.2) and the tsunami caused by it (Lay et al., 2005), which devastated the coastal regions around the Bay of Bengal and killed more than 200,000 people, further exposed the vulnerability of this region due to earthquakes and tsunami. For an objective assessment of the tsunami hazard in this region, it is necessary to understand the tectonics and the mechanism of earthquake occurrence in this region. Other than earthquakes, submarine landslide in the several-kilometer-thick sediments of the Ganga and Brahmaputra rivers deposited in the northern Bay of Bengal may also pose tsunami hazard in the region. There are no historical records to suggest occurrence of such submarine landslides and no studies have been carried out to assess the tsunami hazard due to this. However, some

attempts have been made to assess tsunami hazard due to earthquakes in the region (Okal and Synolakis, 2008; Cummins, 2007).

Tsunami hazard in the Bay of Bengal

The tsunami hazard in the Bay of Bengal region is mainly due to the great and major earthquakes along the northern Sunda arc where the eastern margin of the Indian plate interacts with the Sunda plate (Figure 1). Tsunami caused by great earthquakes in the far-field regions, namely, the Banda arc which is to the east of Sunda arc (i.e., southeast of the equator), the Makran region which is along the southern coast of Pakistan and Iran, and the circum pacific belt, may not affect the Bay of Bengal region due to directivity effect and the presence of intervening land masses during their propagation. The frontal arc of the northern Sunda arc consists of the Indo-Burmese arc in the north, and the Sumatra- Andaman arc in the south with the



Tsunami: Bay of Bengal, Figure 1 Bay of Bengal, general tectonics of the northern Sunda arc and locations of major and great earthquakes along the arc. Stars denote major earthquakes that occurred in the Indo-Burmese arc. These earthquakes were not associated with any tsunami. The great 1897 Shillong Plateau earthquake and the great 1839 earthquake on Sagaing fault are also shown for reference. The 1897 intraplate earthquake is associated with the tectonics of the Indian plate and its interaction with the Himalaya, rather than the tectonics of the Indo-Burmese arc. Major earthquakes of Sagaing fault are not shown here, as they do not pose any tsunami hazard. The 1861 and 1907 earthquakes occurred in the source region of the 2005 Nias earthquake. However, none of the tsunami caused by of these earthquakes, including that of the 2005 Nias earthquake, affected the Bay of Bengal region.

intervening Irrawaddy delta region. In the back arc, the Sagaing fault which runs through Myanmar on the land and is characterized by strike slip motion, the Sumatra Fault system which is characterized by the earthquakes with strike slip motion, and the Andaman Sea ridge transform fault system which is characterized by the normal and strike slip earthquakes of low magnitude ($M < 6.5$), do not pose tsunami hazard to the Bay of Bengal region. The tsunami hazard in the Bay of Bengal due to low-level seismicity in the diffused plate boundary region between India and Australia in the Indian Ocean may be very low.

Plate motion along the northern Sunda arc

In the northern Sunda arc, relative motion between the India and Sunda plates is about 3.7 cm/year toward N15° in the northern part (i.e., Indo-Burmese arc region) and about 4.5 cm/year toward N25° in the southern part (i.e., the Sumatra region). The oblique motion between the India and Sunda plates is accommodated along the arc through slip partitioning in the frontal and back arc (Gahalaut and Gahalaut, 2007; McCaffrey 2009). In the north, the north–south trending Indo-Burmese arc accommodates slip of about 1.6 cm/year through dextral motion (Gahalaut et al., 2010a) and the remaining motion of about 2.0 cm/year is accommodated across the dextral strike slip Sagaing fault in the Myanmar (Maurin et al., 2010). Thus, at present, the underlying gently east dipping Indian plate beneath the accretionary wedge of the Indo-Burmese arc moves toward north. Further south in the Irrawaddy and Andaman region, the motion in the frontal arc becomes oblique. In the back arc, the dextral strike slip motion occurs along the north–south planes and normal motion occurs along the Andaman Sea ridge transform fault system which is consistent with the Andaman sea opening. In the Sumatra region, the motion in the frontal arc is arc normal, leading to thrust motion on the plate boundary interface. The strike slip motion in the back arc is accommodated through dextral motion along the Sumatra Fault System. The Sumatra Fault System joins the Sagaing fault in the north through the Andaman Sea ridge transform fault system.

Seismicity of the northern Sunda arc region and tsunami hazard

For the convenience of discussion, the seismicity of the Indo-Burmese arc, Irrawaddy, and Sumatra-Andaman region is described separately.

Indo-Burmese arc

The plate reconstruction models suggest that subduction probably occurred along the Indo-Burmese arc in the geological past (prior to about 50 Ma) when the arc was predominantly southeast–northwest trending. However, after the collision of the Indian plate with the Eurasian plate, the northern Sunda arc rotated clockwise to become predominantly north–south trending. Several geological and geophysical studies confirm that the subduction occurred across the Indo-Burmese arc between India and Burma plates and there is evidence of the subducted Indian slab, but presently the subduction does not appear to be active (Hall, 1997; Guzman-Speziale and Ni, 1996; Rao and Kumar, 1999; Pesicek et al., 2010). This is further confirmed from the earthquake occurrence and the crustal deformation studies.

The historical records of earthquakes and tsunami in the Indo-Burmese arc are generally very poor. Although there are a few unverifiable reports of earthquake occurrence as early as 1548 in Tripura, Assam (two states in northeastern India), or Bangladesh, there are large uncertainties in the location of historical earthquakes. The catalogue is probably reliable only after 1762. The most notable earthquakes are the April 2, 1762, the August 24, 1858, Arakan earthquakes, and the January 10, 1869, Cachar earthquake (Richter, 1958). Other than these earthquakes, several instances of damage due to earthquakes are reported from Chittagong, Sylhet, Manipur valley, and Cachar regions. However, it is possible that the high damage in these sediment-filled valleys was due to relatively large population and local site effects. The April 2, 1762, Arakan earthquake has been considered as a great tsunamigenic earthquake (Cummins, 2007). It caused extensive damage in the Chittagong region through shaking, liquefaction, damming of channels, seiches, etc. (Gulston, 1763). The reported uplift of Cheduba Island due to this earthquake was not found to be singularly due to this earthquake (Oldham, 1883; Martin and Szeliga, 2010). Halsted (1843) reported the effects of the earthquakes which were mostly found to be highly exaggerated by Gupta and Gahalaut (2009). They found that the highly qualitative observations of the coseismic elevation changes are consistent with a predominant strike slip motion on the earthquake rupture which was located under the land and/or under very shallow water, and it was only a major earthquake (Martin and Szeliga, 2010) which probably did not cause any major tsunami (Gupta and Gahalaut, 2009). The August 24, 1858, Arakan earthquake was felt in many parts of Burma and was severely felt at Kyauk Pyu, Ramree Island, Myanmar. It caused liquefaction and damage to buildings and Pagodas. In this region, the distance between the Sagaing fault and the structurally mapped Arakan trench is less than 200 km, and, hence, based on the scanty reports of damage, it is difficult to judge whether this one and the 1843 and 1848 earthquakes are linked with the Indo-Burmese arc or with the Sagaing fault. But none of these three earthquakes is associated with tsunami. The January 10, 1869, Cachar-Manipur earthquake is referred to as the most severe earthquake in the available 2,000 years of written historical records of Manipur state of India. However, recent examinations of the historical records suggest that even this earthquake was not a great one and the damage was very much confined to the sediment-filled valleys of Imphal and Cachar (Kundu and Gahalaut, 2010b). As this earthquake occurred

on the land, far from the sea, there is no question of tsunami being generated by this earthquake.

Although precise motion on rupture during these major earthquakes is not known, the current seismicity and the results from crustal deformation studies suggest that the historical earthquakes in the Indo-Burmese arc might have occurred through predominant strike slip motion on steep planes, which dips away from the coast. This severely limits the possibility of a tsunamigenic earthquake in the coastal Indo-Burmese arc region. Lack of paleo-tsunami deposits in the coastal Myanmar region (Aung et al., 2008) further supports this argument.

Irrawaddy region

Seismicity in the Irrawaddy frontal arc is extremely low as compared to that in the Andaman Nicobar arc in the south and the Indo-Burmese arc in the north. Earthquakes in the Irrawaddy region occur at depths less than 50 km, whereas in the Indo-Burmese arc and the Andaman Nicobar region, earthquakes occur down to a depth of up to 150–200 km (Guzman-Speziale and Ni, 1996). From the limited earthquake focal mechanisms, it is not possible to constrain sense of motion along this part; however, the regional plate motion models suggest that very oblique motion occurs along the frontal arc (Gahalaut and Gahalaut, 2007). The reason for very low seismicity of the frontal arc at shallow depth (>50 km) could be that the region is either aseismic or accumulating strain. However, absence of seismicity even at depths greater than 50 km suggests that the subducting slab might not be continuing to that depth at all. In the Andaman and Indo-Burmese arc regions, occurrence of earthquakes at intermediate depths confirms the continuity of the Indian plate slab in those regions. The absence of seismicity at intermediate depth level of the Irrawaddy region is possibly due to the presence of tear in the subducting Indian slab, which is has been confirmed from the tomographic studies of the region (Richards et al., 2007). Thus, it is the tear in the slab which is responsible for very low seismicity at shallow depth and no seismicity at deeper depth (Kundu and Gahalaut, 2010a). Presence of a tear and lack of slab in the region may substantially reduce the potential of occurrence of a future great earthquake in the region. Accordingly, it may also reduce tsunami hazard in the nearby regions due to major and great earthquakes in this region. It is possible that this tear hindered the northward rupture propagation of the 2004 Sumatra-Andaman earthquake. However, it is necessary to further confirm the presence of such a tear using marine deep seismic surveys, detailed tomographic studies, and analysis of earthquakes with the deployment of ocean bottom and land seismometers in the region.

Sumatra-Andaman arc region

The state of knowledge about the Sumatra-Andaman region was very poor until the occurrence of the December 26, 2004, Sumatra-Andaman (M 9.2), March 28, 2005, Nias (M 8.6), and several other earthquakes since then. These earthquakes have provided immense information about the tectonics and seismogenesis of this region. These earthquakes suggest that Indian plate subducts obliquely beneath the Andaman arc while the subduction becomes arc normal in the Sumatra arc. Precise hypocentral locations of these earthquakes and their aftershocks have helped in mapping the subsurface plate boundary interface in the frontal arc region and its seismogenic width. The 2004 Sumatra-Andaman earthquake occurred on a 1,400 km long and 100–150 km wide rupture between the northern Sumatra and north Andaman Islands. The large coseismic displacement in the source region (Gahalaut et al., 2006) caused a devastating tsunami which propagated through the Indian Ocean and caused heavy destruction in the countries surrounding the Bay of Bengal, namely, Indonesia, India, Thailand, and Sri Lanka. There were a few remarkable aspects of this earthquake and tsunami caused by it. It is learnt that the rupture in the northern part, i.e., under the Andaman Islands, was slower (Lay et al., 2005) and that it did not reach the surface. Thus this part of the rupture did not contribute in tsunami generation. It has been suggested that the subducting 90°E ridge probably slowed down the rupture under the Andaman Islands (Gahalaut et al., 2010b) and the subducted unconsolidated sediments of Bay of Bengal probably did not allow the rupture to propagate to the surface. These aspects need to be explored further. Earlier, another earthquake occurred here on June 26, 1941, with M 7.7 (Pacheco and Sykes, 1992), for which no seismological data exist to suggest whether slow rupture occurred during that earthquake as well. Though the reports are scanty, it appears that even that earthquake did not generate a tsunami (Rajendran et al., 2007). There may be various reasons for that earthquake for not being tsunamigenic, including slow rupture speed and rupture on blind fault. It may be possible that the Andaman segment may not generate major tsunami during future great earthquakes as well. However, it needs to be explored further by analyzing the frontal arc characteristics through deep seismic imaging and modeling.

An earthquake occurred on December 31, 1881 (M 7.9), in the Nicobar region. It caused a tsunami with less than 1 m height in the Andaman Islands and the coastal region of India. Unfortunately, the history of the Andaman and Nicobar Islands is very poor. The intermittent colonial occupation of the Islands started in the middle of the eighteenth century by the Danish, followed by the British, and continuous colonial occupation started only after 1869 until the independence of India in 1947. Prior to the nineteenth century, we do not know much about the history of the Andaman Nicobar region as the indigenous tribes did not maintain any written records. Thus, prior to the occurrence of earthquakes on October 31, 1847, possibly near Nicobar Islands (Bilham et al., 2005), and the two earthquakes mentioned above, we have no knowledge of historical earthquakes or tsunami in the region. However, it is certain that the great and major earthquakes under the Nicobar and northern Sumatra arc

have produced tsunami and are capable of generating tsunami during future great and major earthquakes.

In the northern Sumatra region, after the occurrence of the 2004 Sumatra-Andaman earthquake, another great earthquake occurred to the south on March 28, 2005, which is referred as the Nias earthquake. Despite its large magnitude and normal rupture speed, the earthquake could not generate a major tsunami. Majority of the slip on the rupture occurred under the shallow and no water and hence it could not displace large amount of water, causing only a localized tsunami whose amplitude faltered quickly as it moved in the open ocean. Nevertheless, this region has the potential to generate a tsunami during great earthquakes, though it may not affect the Bay of Bengal region due to rupture directivity. The arc region located further southeast, though capable of causing tsunami during great earthquakes, may not affect the Bay of Bengal region.

Summary

The Arakan coast of the Indo-Burmese arc can produce a major earthquake. However, it is unlikely that a major tsunami would be generated by the earthquake in this segment due to predominant strike slip motion on the plate boundary interface which is mostly located on the land and under very shallow water, and dips eastward, away from the sea. The Irrawaddy frontal arc region may not produce great or even major earthquakes and hence the tsunami threat from this region is low. The Andaman arc which did not contribute in tsunami generation during the 2004 Sumatra-Andaman earthquake needs to be studied in great detail in view of the slow and blind rupture. The real and definite threat is from the Nicobar and northern Sumatra arc region, which has produced tsunami in the past and is capable of generating tsunami during future great earthquakes.

Bibliography

Aung, T. T., Okamura, Y., Satake, K., Swe, W., Swe, T. L., Saw, H., and Tun, S. T., 2006. Paleoseismological field survey along the western coast of Myanmar. *Annual Report on Active Fault and PaleoEarthquake Researches*, Active Fault Research Center, Geological Survey of Japan, **6**, 171–188.

Aung, T. T., Satake, K., Okamura, Y., Shishikura, M., Swe, W., Saw, H., Swe, T. L., Tun, S. T., and Aung, T., 2008. *Geologic evidence for great Holocene earthquakes off Myanmar*. Phuket, Thailand: International Symposium on the Restoration Program from Giant Earthquakes and Tsunamis, pp. 13–18.

Bilham, R. G., Engdahl, E. R., Feldl, N., and Satyabala, S. P., 2005. Partial and complete rupture of the Indo-Andaman plate boundary 1847–2004. *Seismological Research Letters*, **76**, 299–311.

Cummins, P. R., 2007. The potential for giant tsunamigenic earthquakes in the northern Bay of Bengal. *Nature*, **449**, 75–78.

Gahalaut, V. K., and Gahalaut, K., 2007. Burma plate motion. *Journal of Geophysical Research*, **112**, B10402, doi:10.1029/2007JB004928.

Gahalaut, V. K., Nagrajan, B., Catherine, J. K., and Kumar, S., 2006. Constraints on 2004 Sumatra–Andaman earthquake rupture from GPS measurements in Andaman-Nicobar Islands. *Earth and Planetary Science Letters*, **242**, 365–374, doi:10.10116/j.epsl.2005.11.051.

Gahalaut, V. K., Kundu, B., Singh, S., Kumar, A., Tiwari, R. P., Catherine, J. K., and Ambikapathy, A., 2010a. GPS measurements in the Indo-Burmese arc, (in press).

Gahalaut, V. K., Subrahmanyam, C., Kundu, B., Catherine, J. K., and Ambikapathy, A., 2010b. Slow rupture in Andaman during 2004 Sumatra-Andaman earthquake: consequence of subduction of 90°E ridge. *Geophysical Journal International*, **180**, 1181–1186.

Gulston, E., 1763. An account of an earthquake at Chattigaon: translated from the Persian by Mr. Edward Gulston, in the service of the Honourable East India Company, and communicated by him to the Reverend Mr. Hirst. *Philosophical Transactions*, **53**, 251–256.

Gupta, H., and Gahalaut, V. K., 2009. Is the northern Bay of Bengal tsunamigenic. *Bulletin of the Seismological Society of America*, **99**, 3496–3501.

Guzman-Speziale, M., and Ni, J. F., 1996. Seismicity and active tectonics of the western Sunda Arc. In Yin, A., and Harrison, T. M. (eds.), *The Tectonic Evolution of Asia*. New York: Cambridge University Press, pp. 63–84.

Hall, R., 1997. Cenozoic plate tectonic reconstructions of SE Asia, In Fraser, A. J., Methews, S. J., and Murphy, R. W. (eds.), *Petroleum Geology of Southeast Asia*, Geological Society of London Special publication No 126, pp. 11–23.

Halsted, E. P., 1843. Report on the Island of Chedooba. *Journal of Asiatic Society of Bengal*, **114**, 319–446.

Kundu, B., and Gahalaut, V. K., 2010a. An investigation into the seismic potential of the Irrawaddy region, northern Sunda Arc. *Bulletin of the Seismological Society of America*, **100**, 891–895.

Kundu, B., and Gahalaut, V. K., 2010b. Earthquake occurrence processes in the Indo-Burmese arc and Sagaing fault region. *Tectonics*, (in press).

Lay, T., Kanamori, H., Ammon, C. J., Nettles, M., Ward, S. N., et al., 2005. The great Sumatra-Andaman earthquake of December 26, 2004. *Science*, **308**, 1127–1133.

Martin, S., and Szeliga, W., 2010. A catalog of felt intensity data for 570 earthquakes in India from 1636 to 2009. *Bulletin Seismological Society of America*, **100**, 562–569.

Maurin, T., Masson, F., Claude Rangin, U., Min, T., and Collard, P., 2010. First global positioning system results in northern Myanmar: constant and localized slip rate along the Sagaing fault. *Geology*, **38**, 591–594.

McCaffrey, R., 2009. The tectonic framework of the Sumatran subduction zone. *Annual Review of Earth Planetary Sciences*, **37**, 345–366.

Okal, E. A., and Synolakis, C. E., 2008. Far-field tsunami hazard from mega-thrust earthquakes in the Indian Ocean. *Geophysical Journal International*, **172**, 995–1015.

Oldham, T., 1883. A catalogue of Indian earthquakes. *Memoirs of Geological Survey of India*, **19**, 163–215.

Ortiz, M., and Bilham, R., 2003. Source area and rupture parameters of the 31 December 1881 Mw = 7.9 Car Nicobar earthquake estimated from tsunamis recorded in the Bay of Bengal. *Journal of Geophysical Research*, **108**, 2215.

Pacheco, J. F., and Sykes, L. R., 1992. Seismic moment catalogue of large shallow earthquakes, 1900 to 1989. *Bulletin of the Seismological Society of America*, **82**, 1306–1349.

Pesicek, J. D., Thurber, C. H., Widiyantooro, S., Zhang, H., DeShon, H. R., and Engdahl, E. R., 2010. Sharpening the tomographic image of the subducting slab below Sumatra, the Andaman Islands and Burma. *Geophysical Journal International*, **182**, 433–453.

Rajendran, C. P., Rajendran, K., Anu, R., Earnest, A., Machado, T., Mohan, P. M., and Freymueller, J., 2007. Crustal deformation and seismic history associated with the 2004 Indian Ocean earthquake: a perspective from the Andaman–Nicobar Islands. *Bulletin of the Seismological Society of America*, **97**, S174–S191.

Rao, N. P., and Kumar, M. R., 1999. Evidences for cessation of Indian plate subduction in the Burmese arc region. *Geophysical Research Letters*, **26**, 3149–3152.
Richards, S., Lister, G., and Kennett, B., 2007. A slab in depth: three dimensional geometry and evolution of the Indo-Australian plate. *Geochemistry, Geophysics, Geosystems*, **12**, Q12003, doi:10.1029/2007GC001657.
Richter, C. F., 1958. *Elementary Seismology*. San Francisco: Freeman, p. 768.

Cross-references

Earthquakes and Crustal Deformation
GPS, Data Acquisition and Analysis
Plate Tectonics, Precambrian
Tsunami

TSUNAMI WATCH AND WARNING CENTERS

Shailesh R. Nayak[1], Srinivasa Kumar Tummala[2]
[1]Ministry of Earth Sciences, Government of India, New Delhi, India
[2]Indian National Centre for Ocean Information Services, Hyderabad, India

Synonyms

Maremoto

Definition

Tsunami. Japanese word literally meaning harbor wave (*tsu+nami*). Tsunamis are series of ocean gravity waves caused mainly due to undersea earthquakes and undersea volcanic activity, rarely due to submarine landslides and very infrequently due to meteor impacts.
Tsunamigenic earthquake. Large earthquakes occurring mostly in subduction zones capable of generating tsunami.
Tsunami warning center (TWC). A 24 × 7 operational center with a mission to advise and warn coastal populace on tsunami.
Tsunami Bulletin. A real-time situation report of the event.
National tsunami warning center (NTWC). An early warning system responsible for collating and disseminating tsunami warning advisories information within a country.
Regional tsunami watch provider (RTWP). A NTWC capable of detecting and analyzing tsunami events and informing other NTWC within an oceanic region.
Service level. Level of operational capability in accurately advising/warning target populace.

Introduction

The 2004 Sumatra-Andaman earthquake and subsequent tsunami in the Indian Ocean were a grim reminder of nature's fury. The event prompted many countries in the region to set up tsunami warning centers (TWC), which were further integrated into a collaborative network of regional tsunami watch providers (RTWP) (UNESCO-IOC, 2007). Within India, an interim warning center was established in 2005 (Gupta, 2005; and Nayak and Kumar, 2008a) at Indian National Center for Ocean Information Services (INCOIS) and by the end of 2007, it became a full-fledged operational warning center.

An important fact that is well known about tsunami is that they travel very fast in open ocean (velocity of the order of several hundreds of kilometers per hour) and cannot be discernible easily by wave heights (of the order of a few centimeters), but slow down as they approach the coast (of the order of several tens of kilometers per hour) with wave heights increasing to disastrous proportions (of the order of several meters). Hence, to forewarn ourselves, we need to have a seismic network to detect tsunamigenic earthquakes and a network of sea-level observations to detect any water level changes. A forewarning system is essential to mitigate the effects of tsunami. Few centuries ago, traditional knowledge indicated that as soon the water withdraws from the beach, there is an impending tsunami. However, with the advancement of science (especially modeling, geospatial technologies, and communication infrastructure) and rapid increase in coastal population, the inherent risk in terms of loss of lives and economic damages (socioeconomic parameters) too has increased many fold, but now it is possible to mitigate damage by scientifically confirming/canceling such warnings derived from traditional knowledge and recommending the right advice to the populace at the right time. Such a scientific establishment is known as TWC. The mission of a TWC is to provide early warnings (known as tsunami advisories) on potentially destructive tsunamis to emergency/disaster management officials, and as appropriate, directly to the public to prevent loss of life and property. This article briefly describes TWC, its components, the advisory products that are generated, and level of operational service that these centers are designed for.

Types of tsunamis

Broadly, the Intergovernmental Oceanographic Commission (IOC) of United Nations Educational, Scientific, and Cultural Organization (UNESCO) has classified tsunamis into nine types: Historical tsunami, Local tsunami, Maremoto, Meteorological tsunami (meteotsunami), Microtsunami, Ocean-wide tsunami, Paleotsunami, Regional tsunami, and Teletsunami (Distant tsunami). However, from the early warning perspective, three types are considered most relevant, based on the distance from the source and the tsunami travel time (UNESCO-IOC, 2006). They are:

1. *Local Tsunami* originates from a nearby source for which its destructive effects are confined to coasts within 100 km or less than 1 h tsunami travel time from its source. These can often be the most dangerous because there is often little warning between the causing event and the arrival of the tsunami.

2. *Regional Tsunami* is capable of creating destruction in a particular geographic region, generally within 1,000 km or 1–3 h tsunami travel time from its source.
3. *Distant Tsunami* (also called an ocean-wide, distant, tele or far-field tsunami) is a tsunami that originates from a far away source, which is generally more than 1,000 km away from the area of interest or more than 3 h tsunami travel time from its source. A distant tsunami is capable of causing widespread destruction, not only in the immediate region of its generation, but across an entire ocean. All ocean-wide tsunamis have been generated by major earthquakes in the subduction regions.

Tsunami warning centers in the world

There are seven major tsunami warning centers operating in the world today in terms of independent services. They are: National Oceanic and Atmospheric Administration's (NOAA) Pacific Tsunami Warning Center (PTWC), Japan Meteorological Agency (JMA), NOAA's West Coast and Alaska Tsunami Warning Center, Sakhalin Tsunami Warning Center (STWC), German Indonesian Tsunami Early Warning System (GITEWS), Joint Australian Tsunami Warning Centre (JATWC), and the Indian National Tsunami Early Warning System (INTEWS). Each of these systems is conceptualized as the National Tsunami Warning Centres (NTWC), where they are the sole advisors in terms of area of responsibility regarding tsunami warning. In the event of distant tsunamis, they will act as regional tsunami watch providers (RTWP) to those regions that could be immediately affected and do not have their own NTWC or to other NTWCs operating in the region, as "heads-up" information. In addition to the above, few systems have been conceived as system of systems like International Tsunami Information Center (ITIC) and Indian Ocean Tsunami Warning System (IOTWS), where information from different major tsunami advisory providers is integrated as shown in Figure 1. The latter, IOTWS, has been an initiative from the US and UNESCO to enhance the warning capability in the region. In addition, critical seismic data providers like USGS (Earthworm seismic data processing system) and Global Seismic Network (GSN) are few of the specialized information providers essential for tsunami early warning.

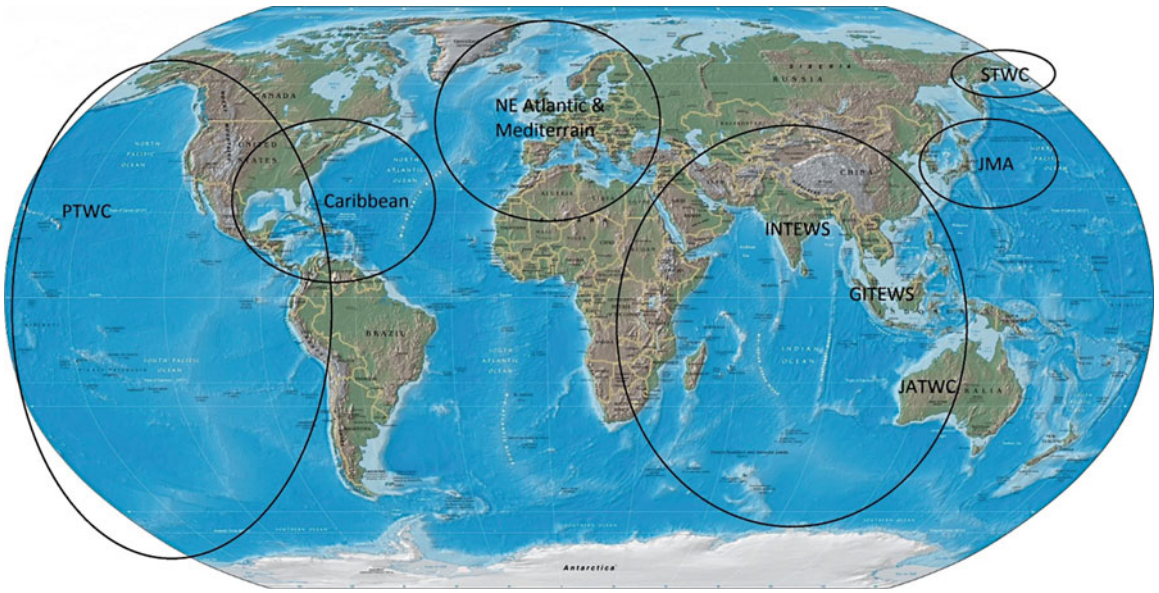
Key components

In order to carry out its mission, the TWC must at least consist of five important components: a network of sensors (seismic and sea level) to detect earthquakes and evaluate its potential for tsunamis, tsunami modeling capability (generation, propagation, and inundation models), high-resolution database on bathymetry and coastal topography, communication infrastructure to issue timely alarms to permit evacuation of coastal areas, and a 24 × 7 operational facility housing the infrastructure, decision support system (DSS) based on standard operating procedures (SOP), and experienced staff (Nayak and Kumar, 2008a). A typical outlay of TWC is shown in Figure 2.

Network of sensors

Seismic network

TWCs use local and global seismographic networks transmitting seismograms in real time to continuously monitor seismicity, in order to locate and estimate the size of potentially tsunamigenic earthquakes. To produce accurate

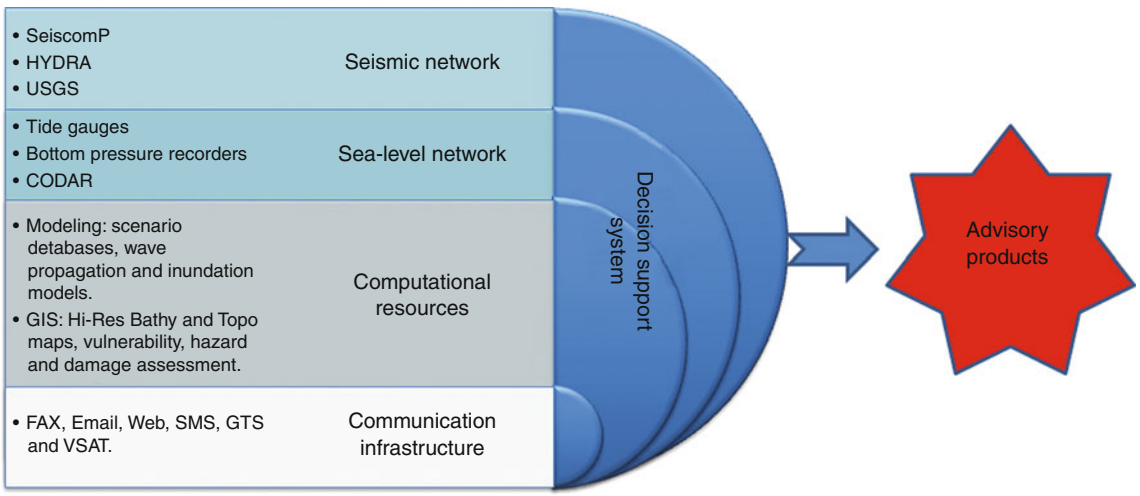


Tsunami Watch and Warning Centers, Figure 1 Tsunami warning systems presently operating in the world.

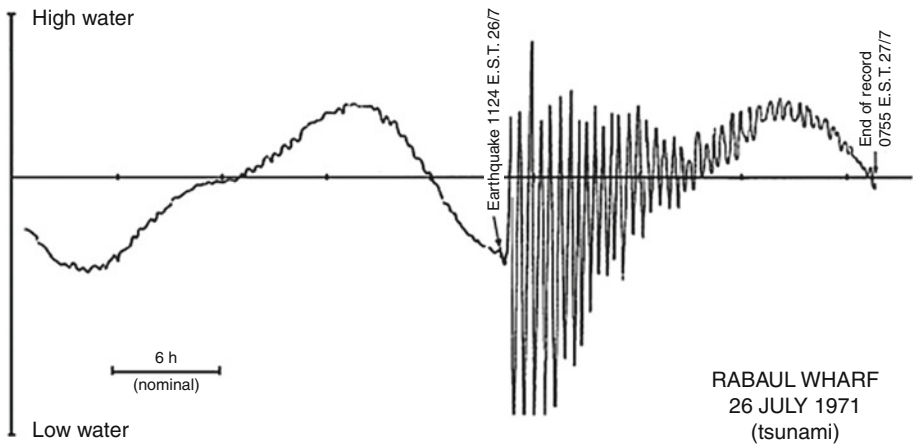
moment magnitude, NTWCs and RTWPs require reliable, broad-frequency, low-noise, high dynamic-range, digital seismic data in real time. The timeliness of the data is crucial in issuing an initial bulletin within 5 min of an earthquake. This is especially important for centers with local tsunami sources. In terms of network density, it is recommended that at least eight stations must be operating in a radius of 900 km, with data latency not exceeding 30 s.

Sea-level network

In order to confirm whether the earthquake has actually caused a tsunami, it is essential to measure the change in water level as near to the fault zone as possible with high accuracy. A TWC uses sea-level networks (tide gauge, bottom pressure recorders, and coastal ocean dynamics applications radar, CODAR) reporting data in real and near real time to verify the generation and evaluate the severity of a tsunami. Numerous international sea-level monitoring networks provide essential real-time data that can be accessed by NTWCs and RTWPs. Many of these networks are coordinated by the IOC of the UNESCO. The most extensive and notable is the Global Sea Level Observing System (GLOSS), conducted under the auspices of the Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) of the World Meteorological Organization (WMO) and the IOC. A tsunami creates a signal component that is not normally present in the observed sea level in usual times. Fortunately, the signature of tsunami waves at sea-level gages



Tsunami Watch and Warning Centers, Figure 2 Typical components of an end-to-end tsunami early warning system. (Courtesy: INCOIS.)



Tsunami Watch and Warning Centers, Figure 3 Typical tsunami signal found in tide gauge sea-level data. (UNESCO Manual of Sea Level Interpretation Vol. 4, 2006.)

is generally quite distinctive and thus recognizable (see Figure 3) and can be differentiated from a signal due to storm surge (UNESCO-IOC Manual and Guides, 2006).

Modeling capability

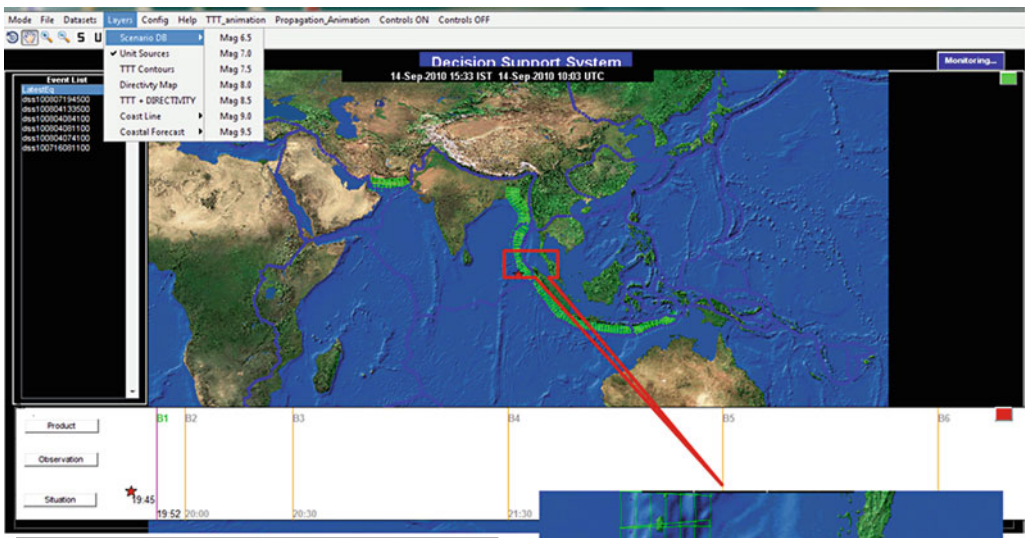
Guesstimating the generation mechanism of tsunami from the earthquake information, identifying the wave propagation scheme, forecasting the travel time, and mapping inundation zones are critical requirements in tsunami early warning, which can be accomplished using numerical models.

Tsunami modeling

Two of the most commonly used tsunami models presently are: the Method of Splitting Tsunami (MOST) developed by Titov and Gonzalez (1997) and the Tohoku University's Numerical Analysis Model (TUNAMI) developed by Imamura et al. (1995). Both these models are capable of simulating the three processes of a tsunami: earthquake generation, transoceanic propagation, and inundation of land. These models basically take the seismic deformation as input to predict the run-up heights and inundation levels at coastal regions for a given tsunamigenic earthquake (Imamura et al., 1995). The seismic deformation for an earthquake is computed using Mansinha and Smylie (1971) formulation using the earthquake parameters like epicenter, focal depth, strike, dip and rake of the rupture plane, length, width of the rupture, and slip magnitude on the rupture plane.

Scenario database

The above models cannot be run in real-time situations because of the computational overheads and resources. Hence, a pre-run scenario database or look-up-table (LUT) approach is most suitable in operational conditions. The forecast strategy is based on a unit-source function methodology, whereby the model runs are individually scaled and combined to produce arbitrary tsunami scenarios. Each unit-source function is equivalent to a tsunami generated by a Mw 7.5 earthquake with a rectangular fault 100 km by 50 km in size and 1 m slip (Gica et al., 2008; Greenslade and Titov, 2008; and Nayak and Kumar, 2008b, c). The faults of the unit functions are placed abutting each other. When the functions are linearly combined, the resultant source function is equivalent to a tsunami generated by a combined fault rupture with assigned slip at each sub-fault. These unit sources are located along the known fault zones for the entire Pacific Basin, Caribbean for the Atlantic region, and Indian Ocean. Figure 4 shows how the unit sources are set up in the Indian Ocean. A database of pre-run scenarios has been created for 1,000 unit sources covering all the tsunamigenic sources in the



Magnitude (M_w)	Width (W) (km)	Length (L) (km)	Slip (u_0) (m)
7.5	50	100	1
8.0	75	150	2.5
8.5	100	300	5
9.0	150	600	10
9.5	300	1200	20

Tsunami Watch and Warning Centers, Figure 4 Modeling using "unit-source" approach and the resulting scenario database. (Courtesy: INCOIS.)

Indian Ocean region (Kuwayama, 2006). At the time of earthquake occurrence, based on the location and magnitude, the basic unit-source open-ocean propagation scenarios are selected from the pre-run scenario database. These scenarios are then merged and scaled up/down using scaling relations (Nayak and Kumar, 2008c) to reconstruct a scenario representing the actual magnitude.

Geospatial integration: inundation modeling and mapping

The resulting open-ocean wave from the scenario database, described in previous section, has to be extended over to the shallow waters and to the coastal area to simulate inundation over land. Knowing the coastal inundation zone is the first step in disaster management and mitigation, as it directly tells us the extent of population to be evacuated and the estimated economic loss due to an event. To perform accurate modeling of tsunami inundation and run-up, a detailed model of near-shore bathymetry and coastal topography is necessary. This is crucial as the accuracy of model predictions is directly related to the quality of the data used to create the bathymetry and topography of the model area. Coastal Bathymetry is the prime determinant of the height of the tsunami wave or storm surge as it approaches the coast. Within the Indian Ocean context, topography of the region at 1:25,000-scale with contours at intervals of 0.5–2 m up to 20 m has been acquired. Preliminary maps have been prepared of coastal topography using satellite stereo data (CARTOSAT-1) for the Indian coast. These datasets, along with airborne **Li**ght **D**etection and **R**anging (LIDAR) data, were further used to create coastal vulnerability index (CVI) maps and multi-hazard vulnerability maps (MHVM) of coastline of India (Nayak and Kumar, 2008b). Similar high-resolution geospatial database is being leveraged in the Atlantic and Pacific Oceans (Gica et al., 2008).

Inundation mapping is not the only area where geospatial technologies are used, the inundation levels estimated and the subsequent threat status has to be accurately linked and conveyed to the respective administrative block within the coastal zones or within a province or a country. To implement such mechanism, the concept of coastal forecast points (CFP) and coastal forecast boxes (CFB) was developed (ICG/IOTWS) and implemented within the Indian Ocean region. Another major application of evolving geospatial trends is the application of 3D modeling or "videogrammetric" approach to visualize inundation and to identify evacuation routes and safe houses during inundation.

Communication infrastructure

Communication infrastructure is the most crucial and tangible aspect of any warning system for tsunami or any hazard. The ability to report and warn of an event is what makes and defines the warning system. Communication in terms of receiving all scientific data and disseminating information to the target audience within the required time needs a reliable communication infrastructure. During this process, TWC must respond fast, be as accurate as possible, and be reliable in order to be effective. All tsunami warning systems feature multiple lines of communications (such as SMS, email, fax, radio, texting, and telex, often using hardened dedicated systems), enabling emergency messages to be sent to the emergency services and armed forces, as well as to population alerting systems.

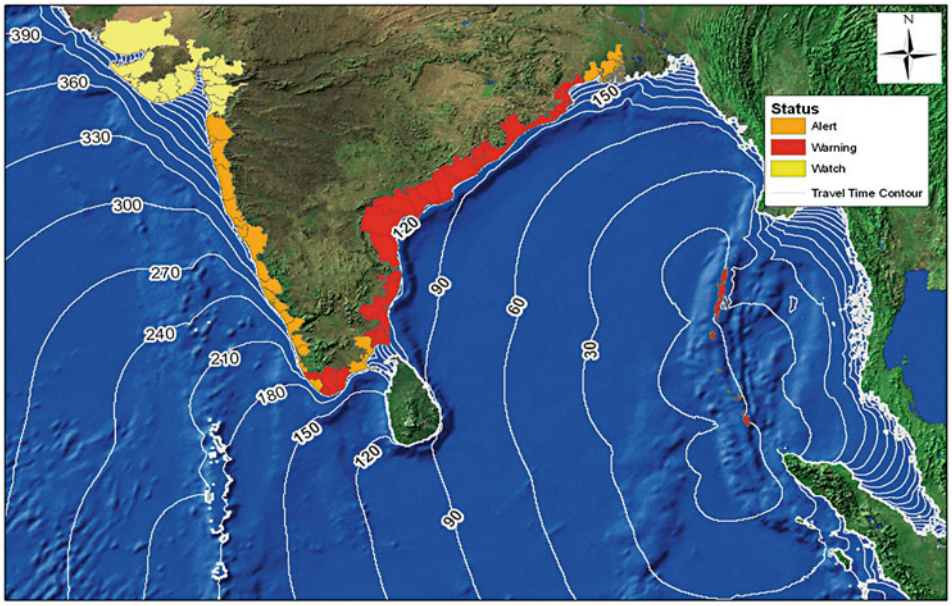
Decision support system and 24 × 7 operational center

In simple words, when the standard operating procedures are coded into a program that can support organizational decision making, then such a program can be called as a decision support system. It helps decision makers to compile useful information from a combination of raw real-time data, documents, personal knowledge, and tsunami scenarios models to accurately disseminate warning advisories to the needed area at the right time. The DSS encapsulates and foresees all the operations from the beginning of the event till the passage of threat by issuing timely advisories (Figure 2). The set of currently practiced advisories are earthquake information, warning, watch, and alert, which are described in detail in the following sections. All the information gathering, decision-making process, and infrastructure are part of the operational monitoring center.

Types of tsunami advisories

Based on the earthquake information, the tsunami modeling scenarios and operational capability, different types of advisory products are issued to the disaster management officials and to the general public. Each advisory has a distinct meaning and clearly suggests the action to be taken. The earthquake information bulletin issued solely on the earthquake information is a qualitative advisory while all others are quantitative advisories (Figure 5) issued taking into consideration the modeling results, and real-time sea-level observations, along with the earthquake information.

1. *Earthquake information bulletin* is issued to inform the occurrence of an earthquake and to advise its potential to generate a tsunami based on the magnitude. The bulletin contains information about the earthquake parameters (origin time, latitude, longitude, focal depth, and magnitude) and the name of the geographical area where it has occurred.
2. *Warning* is the highest level of advisory where immediate action has to be taken. This advisory is issued when a potential tsunami with significant widespread inundation is imminent or expected. Warnings alert the public that widespread, dangerous coastal flooding accompanied by powerful currents is possible and may continue for several hours after arrival of the initial wave. Warnings also alert emergency management



---- TEST TEST TEST TEST -----

TSUNAMI INFORMATION BULLETIN 03
Indian Tsunami Early Warning Centre (ITEWC)
Indian National Centre for Ocean Information Services (INCOIS), Hyderabad
PHONE: 91-40-23895011 FAX: 91-40-23895012 E-MAIL: tsunami@incois.gov.in

Date and Time of Bulletin Issue: 14-Oct-2009 0715 IST

Earthquake Parameters (Revised)

Preferred Magnitude	9.2 Mw (Great)
Network Magnitude(s)	9.2 (mB), 9.2 Mwp, 9.2 Mw(mB), 6.8 (MLv), 6.9 (mb)
Earthquake Origin Date & Time	14-Oct-2009 06:30:00 (IST)
Latitude	3.4 N
Longitude	95.7 E
Focal Depth	20 km
Land/ Ocean	Ocean
Water Level Depth (if Ocean)	3217 m
Region	Northern Sumatra, Indonesia

Tsunami Information

- Sea-Level Observations: Based on Sea level readings, a tsunami has been generated. It may already have been destructive along some coasts.

Tide Gauge Name	Lat	Lon	OWAT (IST)	OWH (m)
Rangat Bay	11.66	92.48	14-Oct-2009 06:47:00	1.96
Campbell Bay	6.90	93.71	14-Oct-2009 06:49:00	2.88
Meulaboh	6.64	93.77	14-Oct-2009 06:35:00	3.63

Legend:
OWAT: Observed Wave Arrival Time (IST)
OWH: Observed Wave Height (m) measured from mean to peak

- Pre-Run Model Scenario: Based on modeled results, the following areas are under WARNING/ALERT/WATCH / NO THREAT

District/Region Name	State/Territory	ETA (IST)	EWH (m)	Status
Hut Bay, Little Andaman	Andaman & Nicobar	14-Oct-2009 06:35	12.75	Warning
North Kamorta	Andaman & Nicobar	14-Oct-2009 06:35	10.23	Warning
West Kamotra	Andaman & Nicobar	14-Oct-2009 06:35	8.08	Warning
Campbell Bay	Andaman & Nicobar	14-Oct-2009 06:35	7.31	Warning
North Car Nicobar	Andaman & Nicobar	14-Oct-2009 06:35	6.39	Warning

Tsunami Watch and Warning Centers, Figure 5 An example of quantitative advisory bulletin product. The product includes graphic of tsunami travel time, along with text indicating the forecast of wave heights and arrival time along with threat status to each administrative zone in the Indian Ocean. (Courtesy: INCOIS.)

officials to take action for the entire tsunami hazard zone. Warnings may be updated, adjusted geographically, downgraded, or canceled.

3. *Watch* is issued to alert emergency management officials and the public of an event which may later impact the watch area. Immediate public evacuation is not required, Local officials should be prepared for evacuation if it is upgraded to warning status. The watch area may be upgraded to a warning or canceled based on updated information and analysis. Watches are normally issued based on seismic information without confirmation that a destructive tsunami is underway.
4. *Alert* contains information about the earthquake and a qualitative tsunami evaluation message indicating that a tsunami is expected. As conditions warrant, the advisory will either be continued, upgraded to a watch or warning, or canceled. Alert implies public should avoid beaches since strong currents are expected. Local officials should be prepared for evacuation if it is upgraded to warning status.
5. *Cancelation* indicates the end of the threat due to tsunami. A cancelation is usually issued after an evaluation of sea-level data confirms that a destructive tsunami will not impact the warned area anymore. A cancelation will also be issued following a destructive tsunami when sea-level readings indicate that the tsunami is below destructive levels and is subsiding in most locations.

Service levels

Based on the level of integration of key component into the TWC, each TWC has the capability to deliver advisory products at certain operational level. Each higher service level represents much more detailed threat evaluation and will be directed to more specific agencies/population group for disaster mitigation with increased confidence levels. Presently, the implementation plan of IOTWS (UNESCO-IOC, 2007) classifies these product levels into three groups, based on incremental capacity development:

Service Level I: At this level of operation, the TWC disseminates only the earthquake information and a qualitative tsunami threat evaluation to other national warning centers.

Service Level II: This is a more robust service level that includes all of the elements of Service Level I but adds modeling and forecast elements so that Watches and Advisory products include: Estimated wave height(s) (EWH) and Estimated time(s) of arrival (ETA) for offshore forecast points, potential threat zones, advanced warning and notification products of threat status and zones in image format. At this service level of operation, real-time observations are included in the decision making and in advisory products from Bulletin – 3 onward.

Service Level III: This is the most advanced level of operational service. At this level, the advisory products include local risk and hazard assessment using inundation models, and they are more specifically directed to communities at risk. Service Level III operations take into account real-time inundation forecast in evaluating the threat status. A TWC operating at this service level indicates a mature early warning system capable of detecting and evaluating tsunamigenic earthquakes occurring anywhere in the global oceans.

Conclusion

Although tsunamis are known to occur rarely, the key elements integrated in a TWC become the essential and minimum requirements in dealing with other ocean-related hazards like: cyclone related storm surges, cloud bursts, and impacts of long-term coastal erosion and sea-level rise. Geospatial technology has immensely helped in the design of early warning system for tsunami, and the GIS base layers and survey data make a TWC capable of dealing with coastal threats due to multiple hazards.

Acknowledgments

The authors would like to thank the Tsunami Early Warning Team at the Indian National Centre for Ocean Information Services (INCOIS) for their support in the preparation of graphics for this manuscript, especially Dr. Uma Devi, Mr. Ch. Patanjali Kumar, and Mr. Raghavendra S. Mupparthy. The authors are thankful to the Director of INCOIS for his active cooperation and encouragement during this process.

Bibliography

Gica, E., Spillane, M. C., Titov, V. V., Chamberlin, C. D., and Newman, J. C., 2008. Development of the forecast propagation database for NOAA's short-term inundation forecast for tsunamis (SIFT). NOAA Technical Memorandum OAR PMEL139. Contribution No. 2937 from NOAA/Pacific Marine Environmental Laboratory.

Greenslade, D. J. M., and Titov, V. V., 2008. A comparison study of two numerical tsunami forecasting systems. *Pure and Applied Geophysics*, **165**, 1991–2001, doi:10.1007/s00024-008-0413-x.

Gupta, H. K. 2005. Mega-Tsunami of 26th December, 2004, 2004: Indian initiative for early warning system and mitigation of oceanogenic hazards. Episodes, March 2005.

Imamura, F., Yalciner, A. C., and Ozyurt, G., 1995. Tsunami modeling manual (TUNAMI model). April (available at: http://www.tsunami.civil.tohoku.ac.jp/hokusai3/E/projects/manual-ver-3.1.pdf)

IOC-ICG/IOTWS-V/13. 2008. Implementation Plan for Regional Tsunami Watch Providers (RTWP), Fifth Session of the Intergovernmental Coordination Group for the Indian Ocean Tsunami Warning and Mitigation System (ICG/IOTWS-V) Kuala Lumpur, Malaysia, 8–10 April 2008.

Kuwayama, T., 2006. Quantitative tsunami forecast system. JICA Tsunami Seminar – Unpublished report.

Mansinha, L. A., and Smylie, D. E., 1971. Displacement fields of inclined faults. *Bulletin of the Seismological Society of America*, **61**, 1433–1440.

Nayak, S., and Kumar, S. T., 2008a. *The First Tsunami Early Warning Center in the Indian Ocean*. Tudor Rose: Risk Wise, pp. 175–177.
Nayak, S., and Kumar, S. T., 2008b. Addressing the risk of tsunami in the Indian Ocean. *Journal of South Asian Disaster Studies*, **1**(1), 45–57.
Nayak, S., and Kumar, S. T., 2008c. Indian tsunami warning system, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII, Part B4, Beijing.
Titov, V. V., and Gonzalez, F. I., 1997. Implementation and testing of the method of splitting tsunami (MOST) model. NOAA Technical Memorandum ERL PMEL-112. Contribution No.1927 from NOAA/Pacific Marine Environmental Laboratory.
UNESCO-IOC. 2006. Tsunami Glossary. IOC Information document No.1221. Paris, UNESCO.
UNESCO-IOC. 2007. IOTWS Implementation Plan for Regional Tsunami Watch Providers, IOC Technical Series 81. Paris, UNESCO.
UNESCO-IOC, Manual and Guides, 2006. Manual on Sea-level Measurements and Interpretation, Volume IV; JCOMM Technical Report No.31; WMO/TD. No.1339: An update to 2006. Paris, Intergovernmental Oceanographic Commission of UNESCO. 78 pp.

Cross-references

Earthquakes, Early and Strong Motion Warning
Tsunami
Tsunami: Bay of Bengal
Seismicity, Subduction Zone