Pure appl. geophys. 163 (2006) 351–372
0033–4553/06/030351–22
DOI 10.1007/s00024-005-0019-5

Pure and Applied Geophysics

The Applicability of Modern Methods of Earthquake Location

Paul G. Richards,[1,2] Felix Waldhauser,[1] David Schaff,[1] and Won-Young Kim[1]

Abstract—We compare traditional methods of seismic event location, based on phase pick data and analysis of events one-at-a-time, with a modern method based on cross-correlation measurements and joint analysis of numerous events. In application to four different regions representing different types of seismicity and monitored with networks of different station density, we present preliminary results indicating what fraction of seismic events may be amenable to analysis with modern methods. The latter can supply locations ten to a hundred times more precise than traditional methods. Since good locations of seismic sources are needed as the starting point for so many user communities, and potentially can be provided due to current improvements in easily-accessible computational capability, we advocate wide-scale application of modern methods in the routine production of bulletins of seismicity. This effort requires access to waveform archives from well-calibrated stations that have long operated at the same location.

Key words: Earthquake location, waveform cross correlation, seismicity studies, California earthquakes, Charlevoix earthquakes, China earthquakes, New Madrid earthquakes.

Introduction

Seismic events are usually still located one-at-a-time by measuring the arrival times of different seismic signals (phase picks) and then interpreting these observations in terms of the travel times predicted for a standard depth-dependent Earth model. In this traditional approach, the differences between observed and calculated arrival times (based on a trial origin time and location) are reduced by a process of iteration (for each event separately) to a value deemed acceptable.

Many different studies of specific regions and particular data sets have demonstrated that by use of whole waveforms and locating groups of events all together, location estimates can be very significantly improved over the results obtained by the traditional approach. In this paper we loosely refer to analysis of waveforms, and joint location of many events, as "modern methods"—in contrast to "traditional methods" based on phase picks and location of events one-at-a-time. We

[1]Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, New York, 10964, USA.
[2]Department of Earth and Environmental Sciences, Columbia University, New York, USA.
E-mail: richards@LDEO.columbia.edu

examine practical aspects of the earthquake location problem for four different regions, and assess the merits of a modern method in which waveform cross correlation is applied to large data sets. We find the fraction of events in each region that are amenable to this type of relocation, and comment on the degree of location improvement.

In regions for which an extensive archive of waveforms can be obtained from stations with a long history of operation at the same site with the same instruments, modern methods of event location (i.e., those that combine waveform cross correlation with a multi-event location algorithm) may now be considered seriously for application to the routine publication of seismicity bulletins.

Sections that follow give background on traditional and modern methods, describe in general our method of analysis for the different regions studied, and present results for the different regions. We conclude with discussion and comment on traditional versus modern methods of event location.

This is not a review paper that compares different methods of cross correlation and event relocation to come up with a verdict about which combination works best in a particular area. The practical experience of network operators will typically be an important guide in determining that combination for different areas. Nor is this an interpretational paper, in which results from each of the four regions are discussed in terms of their implications on the tectonics or earthquake physics. Rather, it is our purpose to present results on relocation using a modern method using waveform cross correlation to the extent permitted by available data and relocating multiple events at the same time; and to comment upon the degree of location improvement, in so far as possible using a common framework of analysis but applied to a broad variety of tectonic regions at different scales, and using seismographic monitoring networks of varying station density.

Background

Seismic data derived from earthquakes and explosions are used in scientific studies of Earth's internal structure, tectonics, and the physics of earthquake processes; in engineering studies of earthquake hazard and efforts in mitigation and emergency management; and in monitoring of nuclear explosions, either as a military activity to evaluate the weapons-development program of a potential adversary, or as an arms-control initiative such as monitoring compliance with the Comprehensive Nuclear-Test-Ban Treaty.

In practice the great majority of those who work with seismic data do not use seismograms directly. Instead they typically use data products derived from seismograms. The most important of these products are bulletins of seismicity, which consist of catalogs of earthquake and explosion locations, measures of event size (such as magnitudes, and scalar and tensor moments), and associated data such as the arrival times of seismic waves at particular stations.

Since the late 1970s there have been enormous improvements in the quality and quantity of seismograms, associated with the deployment of broadband feedback sensors and techniques of digital recording that can faithfully capture signals with high dynamic range across wide bands of frequency; and there is ongoing revolutionary improvement in access to seismogram data, handicapped only by political barriers as reliable satellite communications and the Internet spread even to remote locations and become less expensive. It may therefore seem surprising that routinely published global and regional bulletins of seismicity have not yet seen a corresponding radical improvement that would greatly benefit any of the principal user communities, even though demonstrably successful modern methods of event location and source characterization have been developed and applied in numerous special studies. As discussed further, below, the main reason modern methods have yet to be widely applied is that it is necessary to build up major archives of well-calibrated and easily accessible waveforms from fixed stations operated over many years.

When teleseismic arrival times are used for event location, the resulting location estimates using traditional procedures are typically in error by several km for events detected at hundreds of stations, and by a few tens of km for events detected at tens of stations. Such errors in traditional event locations in standard bulletins are not always appreciated by users, but can be shown for example by relocation studies such as that of ENGDAHL *et al.* (1998), and by comparison of standard location estimates with the hypocenters of events whose ground truth location is known from non-seismic methods (BONDÁR *et al.*, 2004). It can also be demonstrated by study (SCHAFF and RICHARDS, 2004a) of waveform doublets that must be within about 1 km of each other, but that standard bulletins report as having been tens of km apart when the events are located in the usual way (i.e., one-at-a-time from phase picks).

When regional arrival times are used, there is practical experience in the western U.S. to indicate that location uncertainties (as given in conventional bulletins) are at the level of one or two km in areas with high station density (station separation on the order of 10–20 km); and are at the level of about five km in areas with fewer stations (station spacing, approximately 50 km) (personal communication, HAUKSSON, 2005). In broad areas for which events are located with regional signals on the basis of a very sparse set of stations, mislocations can routinely reach several tens of km (RICHARDS *et al.*, 2003).

There are two independent reasons why events can be significantly mislocated, even in situations where station coverage is not a problem. First, there is the difficulty of picking arrival times accurately. When signal-to-noise ratios are good, the error in traditional methods of measuring the arrival time of seismic waves is usually less than one second (and can be less than 0.1 second in favorable conditions). But signals

(*Pn*, for example) are often weak and/or emergent, and later arrivals have to be picked in time windows that include the coda of earlier arrivals. Thus regional *S*-wave phases may be picked with errors that in some cases reach up to a few tens of seconds. Second and often more important, there is the problem of errors in the travel-time model used to interpret measured arrival times. In many cases it is model error, not pick error, which dominates the overall error in absolute location. In such cases the overall goal of improving locations based on seismic arrival times can be achieved only by reducing the effect of model errors. Model errors can also lead to poor-quality relative locations of events that in fact lie near to each other, if the events are located one-at-a-time using data derived from different station sets.

At depths greater than about 200 km, the Earth's laterally-averaged velocity structure is known quite accurately (i.e., to within 1% at most depths, the biggest difference from actual velocities being in regions of subducting tectonic plates). The main difficulty is at shallower depths, within the crust and uppermost mantle, where the actual velocity of seismic waves may differ in unknown ways, perhaps by as much as 10%, from the velocity that is often assumed (such as the velocity given by *ak135* or some other standard Earth model). Thus the problem of substantial differences between actual and standard travel-time models is most significant for regional waves. If a seismic event is about 500 km from a station that detects a regional arrival, and if the arrival is misinterpreted with a velocity that is wrong by 5%, then the event will be estimated from that observation alone as having originated at a distance that is incorrect by about 25 km. Non-horizontal interfaces within the Earth, and phase misidentification (for example between *Pg* and *Pn*) compound these problems.

Experience with the traditional location method originated in the era of analog recording, with sensors of limited dynamic range and bandwidth. It is therefore remarkable that even after decades of experience with digital seismic data beginning in the 1970s, event location is usually still based upon analysis of one-event-at-a-time, and traditional picking of the arrival times for specific seismic phases. In practice this means that only a very small fraction of the available information contained in a modern seismogram is being used to locate events.

In this paper we use a recently-developed highly-efficient waveform cross-correlation algorithm for locally-observed phases (SCHAFF *et al.*, 2004) to measure accurate arrival-time differences for neighboring earthquakes observed at common stations. The waveform cross-correlation method has been applied on the local scale with considerable success. It uses much more of the information contained in seismograms, and can significantly reduce the error in relative arrival times derived from phase picks. We analyze cross-correlation measurements on local and regional signals, combined with ordinary phase picks from earthquake bulletins, using the double-difference technique (WALDHAUSER and ELLSWORTH, 2000; WALDHAUSER, 2001) to obtain precise relative locations for many earthquakes all at the same time.

Several other algorithms have been developed to reduce the effect of errors in the travel time model by locating many events at the same time. These include joint hypocenter determination (JHD) (DOUGLAS, 1967; and FROHLICH, 1979, among others), hypocentroidal decomposition (HDC) (JORDAN and SVERDRUP, 1981), or the use of source specific station terms (SSST) (RICHARDS-DINGER and SHEARER, 2000). All of these methods attempt to correct for correlated effects, due to three-dimensional Earth structure, in the arrival times generated by a cluster of events. The corrections are typically accomplished via use of residual-based station corrections, either in the form of a 1D (static) correction (JHD, HDC), or by using source specific station terms (SSST). The former reduces model errors due to unusual structure beneath a station, while the latter reduces model errors due to structure somewhere between the region of seismicity for which a correction factor is determined, and a particular station. With the double-difference method (DD), a network of events is built in which each event is linked to its nearest neighbors through travel-time differences observed at common stations, which are directly inverted for event separation. The method thus differs from earlier methods of joint hypocentral determination in that no station corrections are necessary, because unmodeled velocity structure is directly removed for each event pair individually. Thus double-difference results are best if events are densely distributed (a few km event separations when using local stations separated by several tens of km). For studies based on regional signals (distances typically in the range from a few hundred km up to about a thousand km), events can be more sparse although event pairs need to have approximately the same take-off direction for corresponding wave types recorded at any common station. The web of DD-relocated events can be expected to give precise relative locations for neighboring events, particularly when pick error is avoided by use of cross correlation. It is not our purpose here to compare these different algorithms, which we would expect to give similar results for clusters of events. Waveform cross-correlation methods together with multi-event location algorithms have been applied by many previous investigators (for example POUPINET *et al.*, 1984; ISRAELSON, 1990; HARRIS, 1991; GOT *et al.*, 1994; WALDHAUSER *et al.*, 1999 and 2004a,b; RUBIN *et al.*, 1999; WITHERS *et al.*, 1999; THURBER *et al.*, 2001; SCHAFF *et al.*, 2002; ROWE *et al.*, 2002; XIE *et al.*, 1997, and XIE, 2001; SHEARER, 1997, HAUKSSON and SHEARER, 2005, and SHEARER *et al.*, 2005). Most of these studies focused on local earthquakes, though a few used regional and teleseismic waveforms to locate small clusters of events. These papers have convincingly demonstrated major improvements in the precision of relative (and in some cases absolute) event locations. In the present paper we are interested in evaluating a modern method (i.e., waveform cross correlation to the extent permitted by available data, combined with multiple event location) on different scales, including some examples at quite large scales; and in commenting generally on the advantages for the modern approach that accrue at larger scales.

Method of Analysis for the Different Regions

For each of the four regions of seismicity to which we have applied modern methods of event location, we have taken the three steps of (a) data collection, (b) waveform analysis, and (c) event relocation using the double difference (DD) algorithm.

The four regions of seismicity for which we have obtained relocations and on which we comment in this paper are:

(1) A broad-area study of the seismicity of China and surrounding regions from 1985 to 2000, described in more detail by SCHAFF and RICHARDS (2004a). This study built upon a relatively small study of foreshocks, main shock, and aftershocks, over a period of several months for the Xiuyan region of Liaoning Province, China, described in more detail by SCHAFF and RICHARDS (2004b). The small study used regional seismograms in the distance range 600 to 1500 km archived by the IRIS Consortium's Data Management Center (IRIS DMC), and showed that complex *Lg* waveforms correlated well for a significant fraction of events for time windows of several hundred seconds duration. The larger study of China was based on 115 stations at distances up to 20° from about 14,000 events, with measurements made on about 130,000 seismograms. Cross correlation was based on window lengths from 60 to 400 s beginning prior to the first *P*-arrival, ending 40 s after *Lg*, and signals passed in the band from 0.5 to 5 Hz. 1.2 million cross correlations were made, for signals associated with event pairs that (according to the Annual Bulletin of Chinese Earthquakes) were not more than 150 km apart.

(2) A study of the New Madrid region of the central United States, for two different eras of station deployment. Data consisted of phase picks as well as waveforms (used for cross-correlation measurement), associated with seismicity bulletins of events located by traditional methods applied by local network operators. These were the Center for Earthquake Research Information of the University of Memphis, which deployed a PANDA network (42 stations, 918 events, 17,598 phases) from October 1989 to August 1992, and a different network (85 stations, 614 events, 16,461 phase picks) from January 2000 to October 2003. Window lengths of 1 s were used for cross-correlation measurements.

(3) A study of the Charlevoix region of Eastern Canada from January 1988 to December 2003 involving 2797 events recorded at 46 stations operated by the Geological Survey of Canada with 33,423 phase picks. Window lengths of 1 s were used for cross-correlation measurements. In practice, almost all of the highly cross-correlated seismograms derived from only 7 stations, closest to the Charlevoix region.

(4) A study of the region monitored by the Northern California Seismic Network (NCSN). This was by far the largest study, with about 225,000 events recorded by 900 stations from 1984 to 2003. Window lengths of only 1 and 2 seconds were used, and the NCSN seismicity bulletin (prepared using traditional methods) was used to select these time windows for all events whose inter-event distance was up to 5 km. Further details are given by SCHAFF and WALDHAUSER (2005).

The associated waveforms for sets of previously defined events were obtained for study (1) from the IRIS DMC; and for study (4) from the Northern California Earthquake Data Center (including phase picks in this case). Though large amounts of data were requested, these two acquisitions were straightforward in that both these data centers had archives previously prepared with the expectation that outside scientists would be using them. But for studies (2) and (3) the basic data sets of associated waveforms and phase picks entailed considerable effort to assemble (with the willing help of network operators). It was necessary to account for different formats, different station sets, and different practices of bulletin preparation over the periods of time during which the seismicity in these regions had been documented.

For some purposes, such as scientific study of earthquake interactions in a fault zone or seismic sources associated with magma conduits in a volcano, relative locations can be sufficient. But of course for seismic monitoring of nuclear explosions, or identification of earthquakes with particular fault structures, there is the need to move beyond relative locations and to estimate absolute hypocentral coordinates. In this paper we focus on improvements in precision, i.e., on relative locations, noting that absolute locations may be achievable at a later step in several different ways—for example by use of ground truth information for a subset of events, and/or by use of a good 3-D travel-time model.

Preliminary Results for Four Different Regions

(1) China and Surrounding Regions of East Asia

This project began with a small preliminary study of 28 earthquakes drawn from a sequence of 90 events during 1999–2000 in the Xiuyan region of Liaoning Province, and moved on to an analysis of about 14,000 events for all of China.

The small study surprisingly showed that in some cases the complex, highly scattered *Lg*-wave is similar at far-regional distances for clusters of events (Schaff and Richards, 2004b). Figure 1 shows the high degree of similarity for more than 300 s of waveform, even at frequencies up to 5 Hz. Cross correlations provided highly accurate differential travel-time measurements. Their error estimated from the internal consistency is about 7 ms. These travel-time differences were then inverted by the double-difference technique to obtain epicenter estimates that have location precision on the order of 150 meters, as shown in Figure 2. The locations were computed with waveform data acquired by four or five regional stations 500 to 1000 km away. The relative epicenter estimates were not substantially affected by the paucity of stations or by large azimuthal gaps. For example, using four stations with an azimuthal gap of 133°, or using only three stations with an azimuthal gap of 220°, resulted in locations that differed by less than 150 m.

Regional event locations must often be based on a small number of phases and stations due to weak signal-to-noise ratios and sparse station coverage. This is

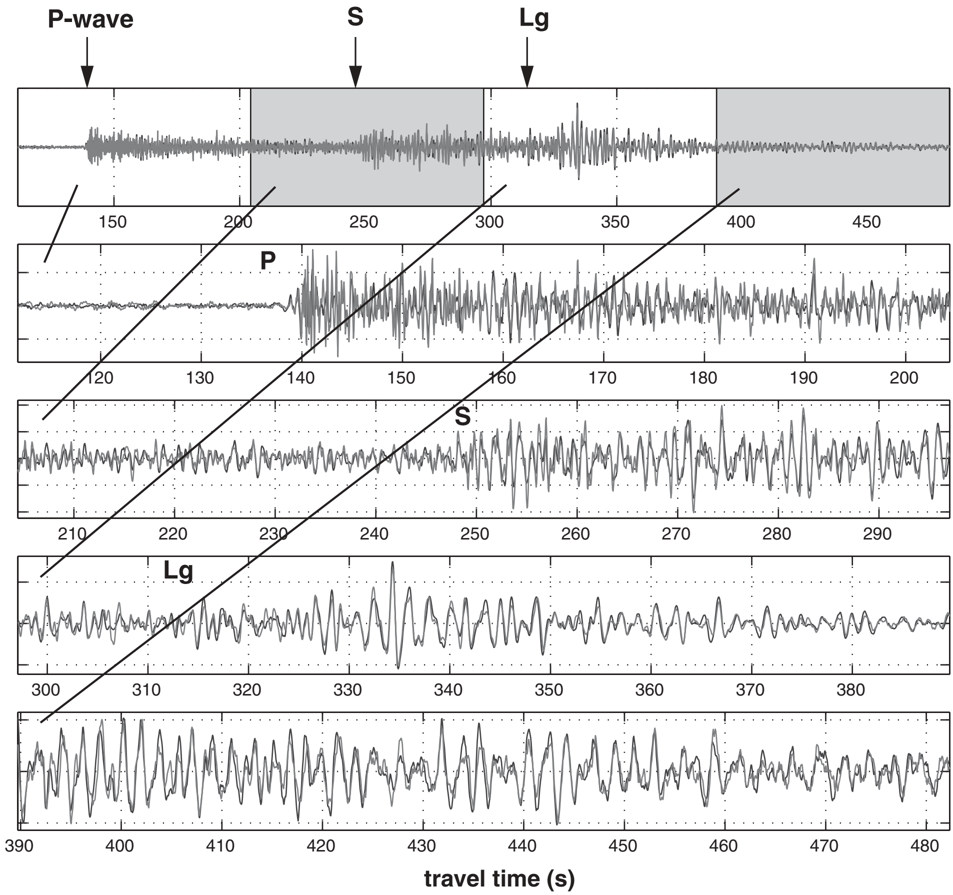

Figure 1
A pair of similar events in China filtered from 0.5 to 5 Hz and superimposed in gray and black. Vertical axes are normalized to unit amplitude. Lower subpanels are enlargements of the white and gray segments. The predicted *P* wave arrives at 143 s, the *S* wave at 256 s, and the *Lg* wave at 315 s. It is apparent that these waveforms are very similar. The quality of this waveform doublet is typical of the events described in more detail by SCHAFF and RICHARDS (2004ab), who argue that such events must be not more than about 1 km apart.

especially true for monitoring work that seeks to locate smaller magnitude seismic events with a handful of regional stations. Two primary advantages of using *Lg* for detection and location are that it is commonly the largest amplitude regional wave (enabling detection of smaller events); and it propagates more slowly than *P* waves or *Sn* (resulting in smaller uncertainty in distance, for a given uncertainty in travel time). This preliminary study demanded a high standard for identification of similar events (cross correlation ≥ 0.8 for a several hundred sec window of signal passed in the band from 0.5 to 5 Hz).

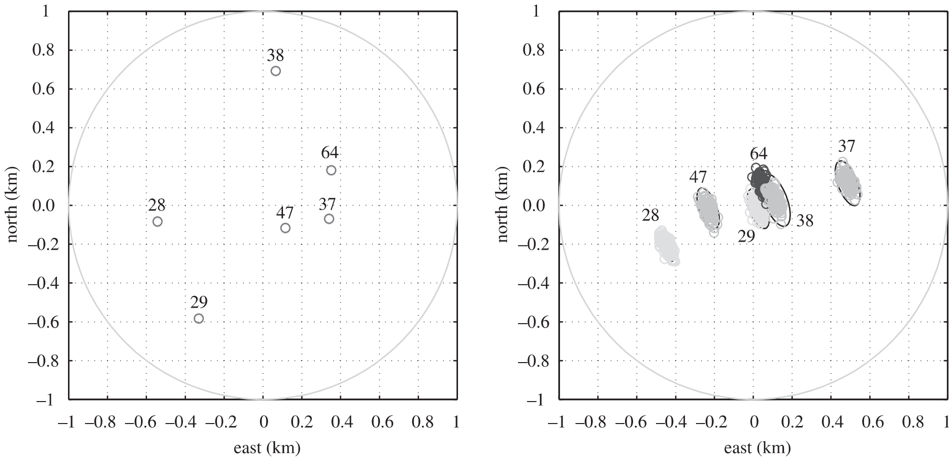

Figure 2
Comparison of double-difference relative locations for a subset of events in the sequence for a local/regional network (left) using only *P*-wave phase picks recorded at several hundred stations and for a sparse regional network (archived at IRIS) using *Lg* cross-correlation measurements (right). Event numbers are for identification. The RMS travel-time residuals are about 1 sec for the *P* waves and 0.02 sec for *Lg*. 95% confidence formal error ellipses and bootstrap errors (shaded small circles) are in good agreement (right). The epicenter in each case is taken as the centroid of the cluster. For further details see SCHAFF and RICHARDS (2004b).

In the larger study we found that about 9% of seismic events in and near China, from 1985 to 2000, were repeating events not more than about 1 kilometer from each other (SCHAFF and RICHARDS, 2004a). This conclusion was based on the stringent criterion for waveform similarity described in Figure 3 (see caption). We cross-correlated seismograms from about 14,000 earthquakes and explosions and measured relative arrival times to within about 0.01 seconds, enabling lateral location precision of about 100–300 meters. Recognition and measurement of repeating signals in archived data and the resulting improved locations enabled us to quantify the inaccuracy of current procedures for picking onset times and locating events. The fraction of cross-correlated events, though quite low (9%) in Figure 3 for the entire time period from 1985 to 2000, rose to a larger value (14%) in the later years, as shown in Figure 4. Table 1 shows too that the fraction increased significantly if the stringent criterion ($CC \geq 0.8$) is relaxed. Thus, relaxation to $CC \geq 0.6$ resulted in 23% of the events in the ABCE cross correlating successfully for the whole time period.

(2) New Madrid, Central United States

This region of intraplate seismicity in the Central United States experiences about 200–250 locatable earthquakes each year. It has been monitored by various networks since the 1970s, and currently we have preliminary relocation results for one three-year period and one four-year period.

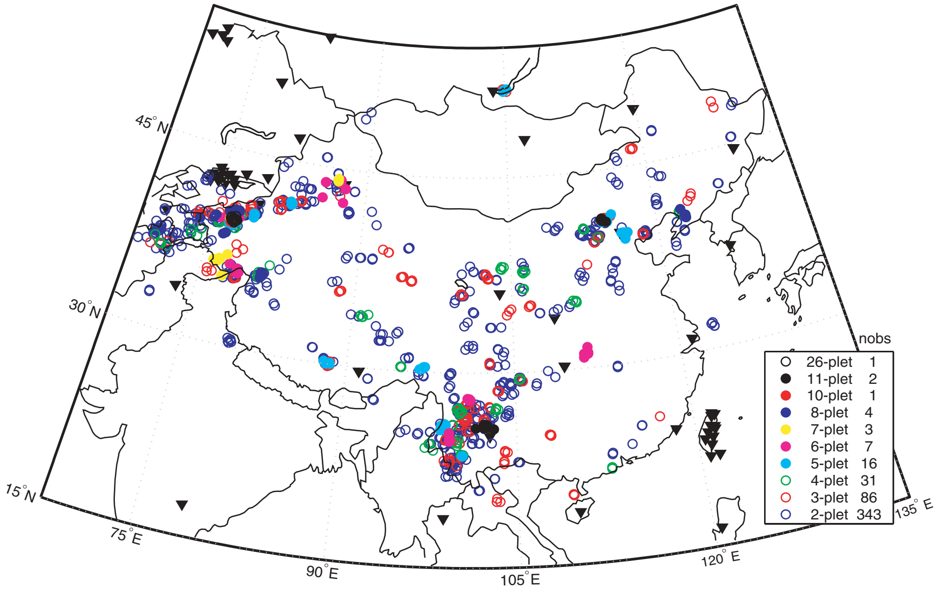

Figure 3
1301 events (9% of the Annual Bulletin of Chinese Earthquakes – ABCE), whose seismograms satisfy the criterion of cross-correlation coefficients greater than or equal to 0.8 with seismograms from at least one other earthquake, for long windows from 5 seconds before the *P* wave to 40 sec after the *Lg* wave on waveforms that are filtered from 0.5 to 5 Hz (as shown in Fig. 1). There are 494 multiplets here, and the inset shows the number of multiplets containing between 2 and 26 events. Recording stations archived by the IRIS Consortium are denoted with solid black triangles. Events are plotted at their ABCE absolute locations. For further details see SCHAFF and RICHARDS (2004a).

Thus, Figure 5 shows relocation of 783 events monitored by a 42-station network from 1989 to 1992. Although 707 of the events (90%) had five or more *P*-wave cross correlations (CC ≥ 0.7), these were not necessarily to the same event. It is possible that these events are relocated with high precision, but this is not assured. We found that a subset of 616 events (65%) had five or more *P*-wave cross correlations (CC ≥ 0.7) to a neighboring event and were thus relocated with high precision provided the azimuthal distribution of stations was adequate; 695 events (85%) cross-correlated at four or more stations; and 735 (90%) at three or more stations. Figure 6 shows relocation of 594 events that occurred in the period January 2000 to October 2003 and were monitored by an 82-station network still operating in 2005. In this case, 499 events (84%) had five or more *P*-wave cross correlations (CC ≥ 0.7), but these were not necessarily to the same event. 371 events (63%) had five or more *P*-wave cross correlations (CC ≥ 0.7) to a neighboring event and were thus relocated with highest precision in our study. Events that do not cross correlate are still located more precisely than in the traditional bulletin, because of the use of the double-difference algorithm applied to phase-pick data for the whole set of events.

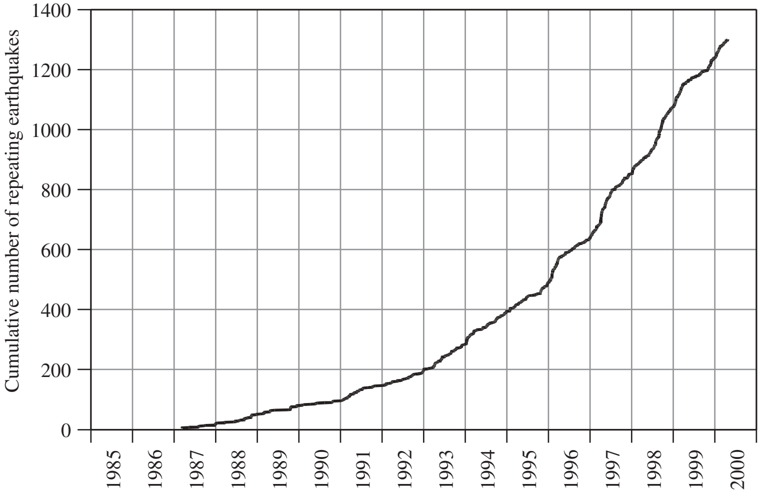

Figure 4
Cumulative number of events vs. time for the 1301 repeating events in China shown in Figure 3 for the period from 1985 to early 2000. Network coverage is more sparse for earlier years, resulting in underestimation of the actual percentage of repeating events over the entire period. If only the 908 repeating events later than 1994 are considered (out of the total of about 6400 events reported in the Annual Bulletin of Chinese Earthquakes), then about 14% of the events satisfy the stringent criterion of cross correlation not less than 0.8 for a window from the *P* arrival until 40 s into the *Lg* signal.

The general distribution of seismicity is very similar between Figures 5 and 6. Figure 7 shows the locations obtained by traditional methods for the same period and station set used in Figure 6. It is apparent especially in the cross sections and in the map view of events in the southwest subregion, that the relocated events in Figure 6 more clearly identify lineations than do the traditionally-obtained locations of Figure 7.

Table 1

The percentage of events in the Annual Bulletin of Chinese Earthquakes (ABCE) is shown for different values of the cross-correlation value. The effect of relaxing the stringent criterion ($CC \geq 0.8$) is substantial. (The analysis was done for 14,000 ABCE events, 130,000 seismograms, 1.2 million CC.)

cross-correlation value (CC)	% of events in ABCE
0.9	3
0.8	9
0.7	16
0.6	23
0.5	30
0.4	38
0.3	52
0.2	69
0.1	76

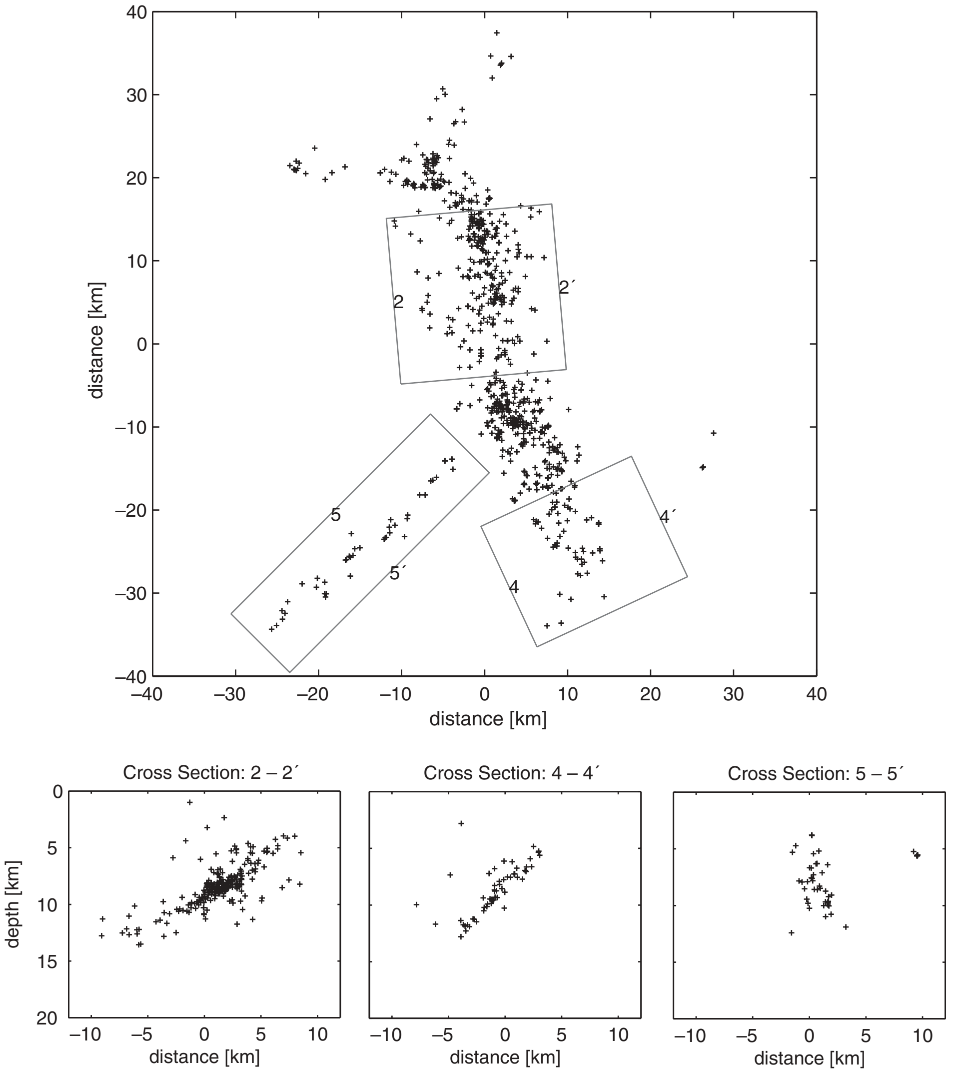

Figure 5

Map view and three cross sections for 783 relocated seismic events in the New Madrid region of the central United States from October 1989 to August 1992, using 42 stations of a PANDA network (see Chiu *et al.*, 1992). This application of the double-difference algorithm was based upon 58,807 phase-pick pairs and 80,697 cross correlations (CC $\geq$ 0.7) derived from *P* waves, plus 66,581 phase-pick pairs and 81,750 cross correlations (CC $\geq$ 0.7) derived from *S* waves. Cross sections show clear evidence for a westward-dipping fault plane, the dip increasing from about 30° (section 2–2′) to 45° further south (section 4–4′).

(3) Charlevoix, Eastern Canada

This too is a region of intraplate seismicity. The preliminary relocations shown in Figure 8, as well as the original bulletin locations for this region, indicate that

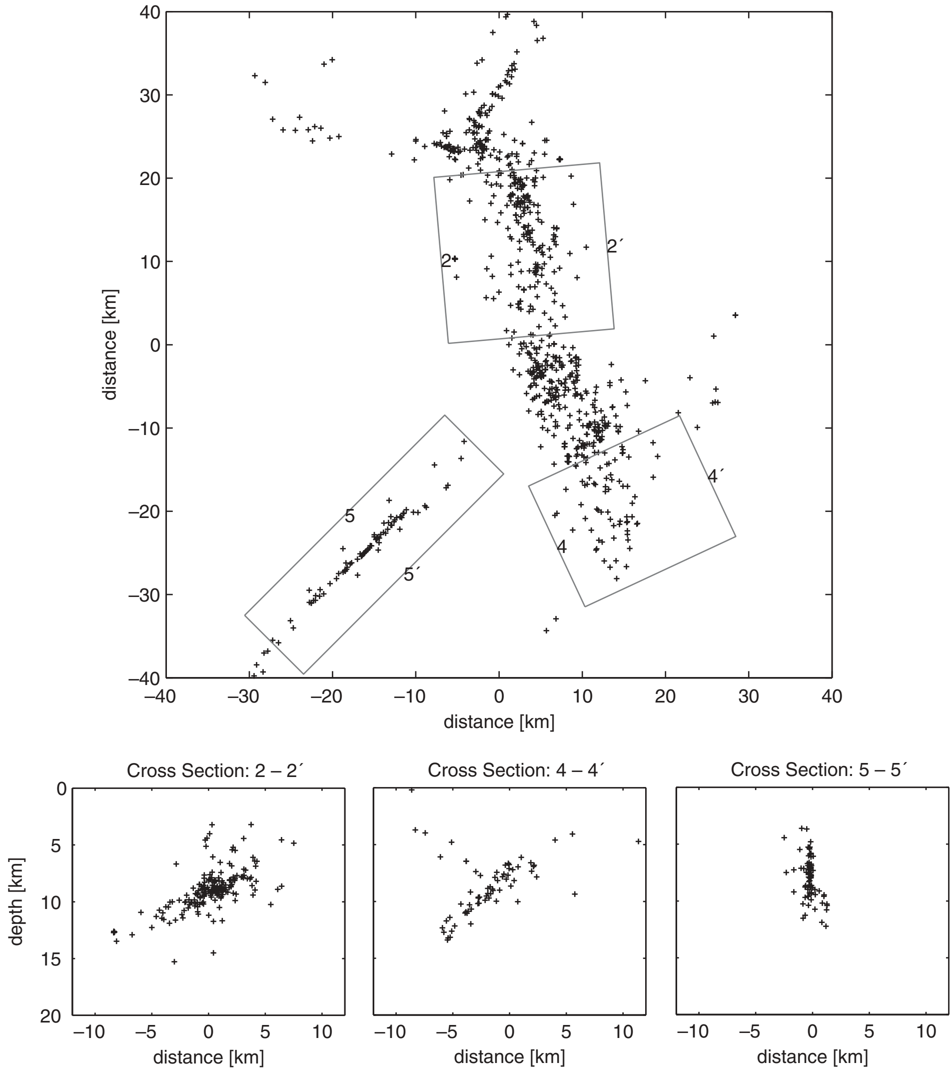

Figure 6
Map view and three cross sections for 594 relocated seismic events in the New Madrid region from January 2000 to October 2003, using 85 stations operated by the University of Memphis. This application of the double-difference algorithm was based upon 202,249 phase-pick pairs and 49,660 cross correlations (CC ≥ 0.7) derived from *P* waves, plus 192,277 phase-pick pairs and 93,009 cross correlations (CC ≥ 0.7) derived from *S* waves.

active faulting is simpler and more clearly defined in the northeastern part of the Charlevoix seismic zone, compared to more complex features in the southwestern part. The dominant faults are two parallel southwest-northeast running structures that dip to the southeast. Of these, the more eastern fault has a dip of about 50°

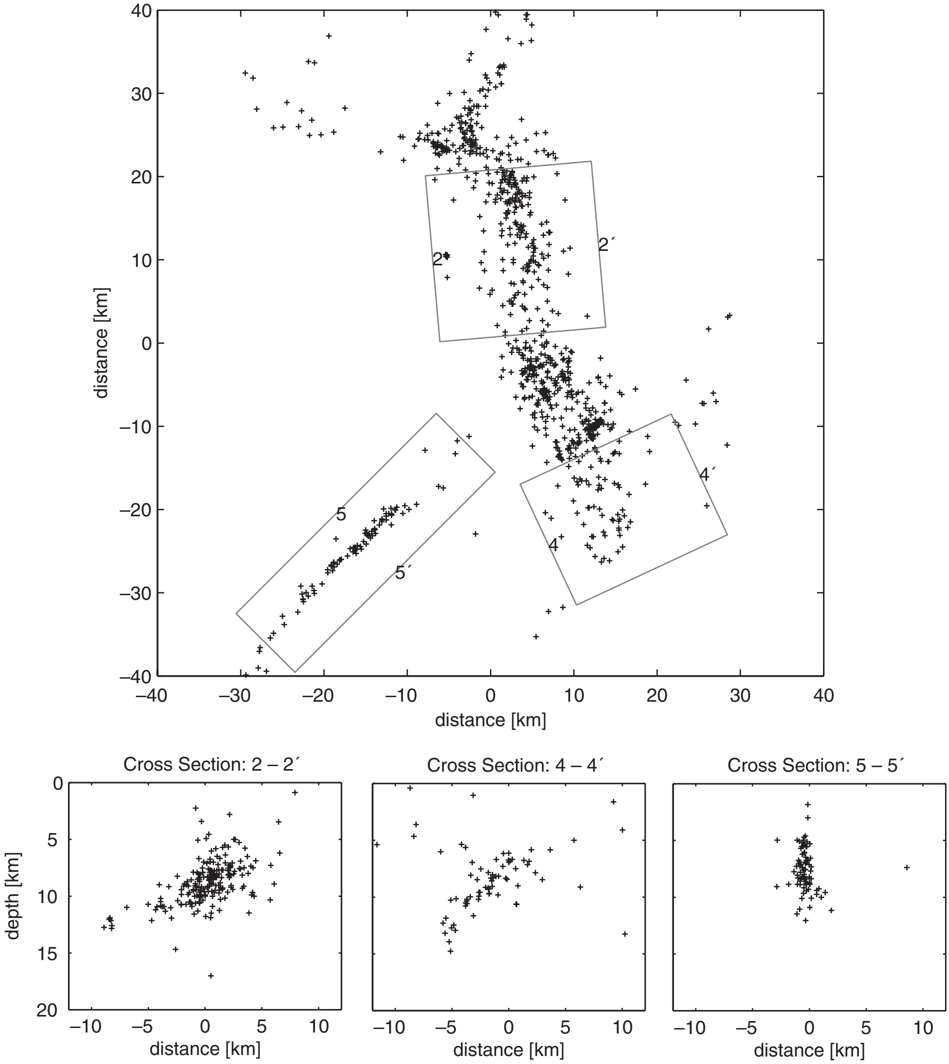

Figure 7
The traditionally-obtained event locations (catalog locations), in map view and for three cross sections, for the station set and time period of Figure 6.

(cross section 2–2′) and that to the northwest has a dip of about 75° (cross section 3–3′). In the northeast the seismicity starts at about 7 or 8 km depth; in the southwest it starts at about 4 km depth. Seismicity extends down to about 30–35 km depth. Larger events occur in the northeastern sub region (circled events have magnitude ≥ 4).

For the Charlevoix seismicity shown as 2272 relocated events in Figure 8, only 242 of them (10%) cross-correlated (with CC ≥ 0.7) at 5 or more stations; 622 (25%)

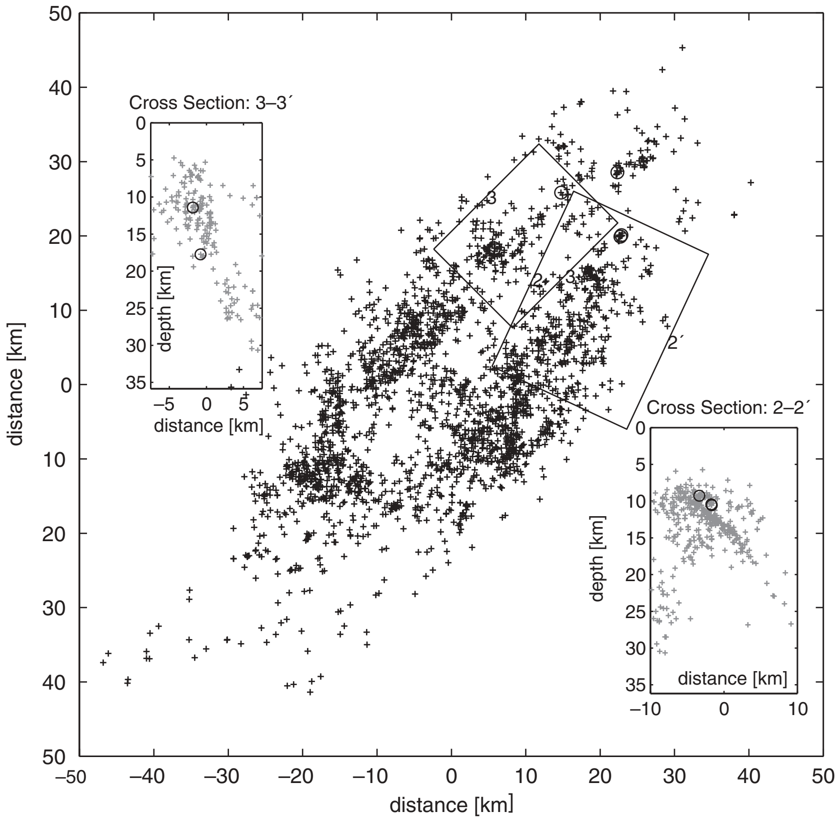

Figure 8
Relocation of 2272 events in the Charlevoix region of Eastern Canada from January 1988 to December 2003, using 46 stations operated by the Geological Survey of Canada. This application of the double-difference algorithm was based upon 495,347 phase-pick pairs and 127,600 cross correlations (CC ≥ 0.7) derived from *P* waves, plus 567,823 phase-pick pairs and 153,510 cross correlations (CC ≥ 0.7) derived from *S* waves. Circled events have magnitude ≥ 4.

at 4 or more stations; and 1439 (57%) at 3 or more stations. Nevertheless, the double-difference algorithm applied to a data set that was largely comprised of phase-pick differences rather than cross-correlation measurements still resulted in improved locations. Figure 9 compares a cross section of the seismicity located by the traditional method and by modern methods (phase picks plus cross correlation, and double difference). The latter cross section more clearly shows a lineation that presumably indicates faulting.

(4) Northern California

Schaff and Waldhauser (2005) describe results from an application of cross-correlation methods to process the complete digital seismogram database for northern

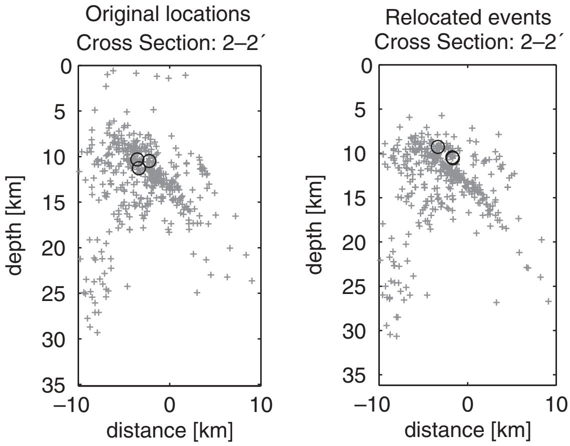

Figure 9
For the Charlevoix cross section 2–2′ of Figure 8, here are shown the locations obtained by traditional methods (on the left) and the relocations using phase picks plus a relatively small number of relative arrival times measured via cross correlation, and relocating events all together. The relocated events more clearly delineate a linear structure, dipping about 45° to the southeast.

California to measure accurate differential times for correlated earthquakes observed at common stations. The waveform database includes about 15 million seismograms from about 210,000 local earthquakes between 1984 and 2003. A total of 26 billion cross-correlation measurements were performed on a 32-node (64 processor) Linux cluster. All event pairs with separation distances of 5 km or less were processed at all stations that recorded the pair. A total of about 1.7 billion *P*-wave differential times had cross-correlation coefficients (CC) of 0.6 or larger. The *P*-wave differential times are often on the order of a factor of ten to a hundred times more accurate than those obtained from routinely picked phase onsets. 1.2 billion *S*-wave differential times were measured with CC $>$ 0.6, a phase not routinely picked at the Northern California Seismic Network because of the generally weak onset of *S* phases, often obscured by *P*-wave coda. These results show a surprisingly high degree of waveform similarity for most of the Northern California catalog, which is very encouraging for improving earthquake locations. Overall statistics are that for each of about 200,000 events (95% of the total), waveforms have CC values that are greater than 0.7 for at least four stations with one or more other events. 90% of the events meet this criterion at eight or more stations, and 82% of the events in the catalog cross correlate at twelve or more stations. To illustrate the spatial distribution of correlated events, Figure 10 shows the percentage of events, within bins of 5 km $\times$ 5 km, that have CC $>$ 0.7 for *P*-wave trains with at least one other event at four or more stations. Even tectonically complicated zones exhibit favorable statistics, such as Long Valley Caldera and Geysers Geothermal Field, where mechanisms are quite variable. Apparently, as long as the earthquake density is high enough there is a high probability that at least one other

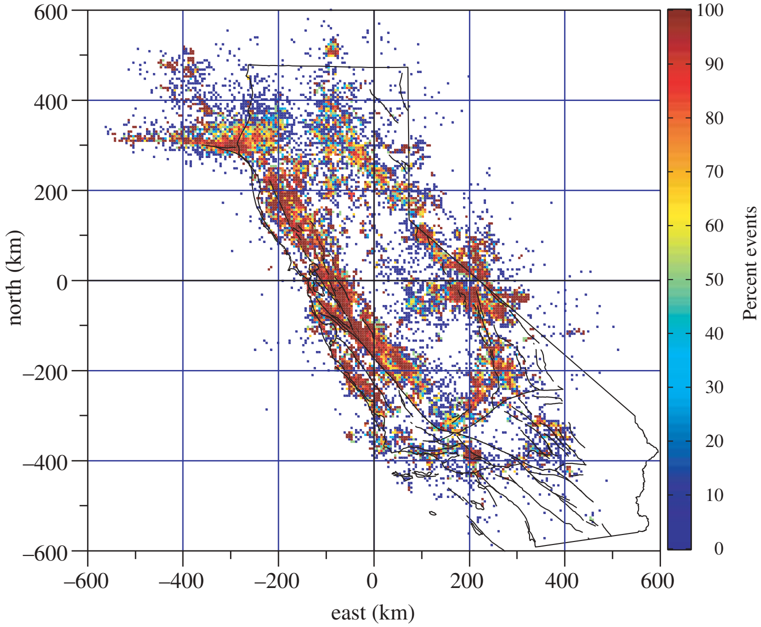

Figure 10
There are more than 200,000 events in the Northern California catalog since 1984 for which waveforms cross correlate (at CC > 0.7) with at least one other event at four or more stations. Here is shown the percentage of events in 5 km by 5 km bins that meet this criterion.

event occurs nearby with a similar focal mechanism, enabling very precise relative locations. To first order, the areas of highly correlated events agree with areas of dense seismicity, suggesting that the closeness of events is the main factor for producing similar waveforms. In a preliminary relocation of 80,000 of these Northern California events, WALDHAUSER *et al.* (2004b) obtained results exhibiting detailed features, such as streaks of seismicity, very similar to those obtained in previous studies of event subsets such as the Calaveras fault seismicity.

Discussion and Conclusions

We have assembled and analyzed waveform data sets and phase-pick data sets to assess the capability of modern methods of seismic event location, as compared to traditional methods, for studies of seismicity in China and North America. After making waveform cross-correlation (WCC) measurements within each station archive and combining them with phase-pick differences, the double-difference (DD) relocation algorithm has been applied for each of the study areas. The resulting location estimates are significantly improved over traditional estimates. In some cases, the improvement is very significant and is associated with a high percentage of events that cross correlate. But even without such CC measurements, there is improvement.

Many other authors have carried out related studies. Thus, RUBIN *et al.* (1999) applied modern methods to relocate about 75% of the events on a creeping section of the San Andreas fault, California. PENG and BEN-ZION (2005) have reported their analysis of about 18,000 aftershocks in the source regions (approximately 80 km by 40 km) of the 1999 Izmit and Duzce earthquakes in North Anatolia, Turkey. The events had been located quite accurately using traditional methods, and though these authors did not relocate the events they cross correlated signals at common stations of a ten-station PASSCAL deployment for those events estimated to be not more than 10 km apart. They found that about 60% of the events formed 2000 similar-event clusters (442 of which had five or more events).

For the four regions we examined in this paper, China was studied with far-regional signals and the other three regions were studied with local stations. Yet we found significant differences in the percentage of improved relocations even among the latter three. Specifically we found that about 85% of New Madrid events could be relocated with modern methods (using waveform cross correlation), and 95% of Northern California events; but only about 25% of Charlevoix events. It is interesting to find that New Madrid and Northern California have approximately the same station density (about one station per 100 km^2) and distance between stations (about 12 km) whereas Charlevoix is a factor of two smaller in density and a factor of two greater in station distance. Also, we found that primarily seven of the Charlevoix stations provided almost all the useful cross-correlation measurements. The other stations of the network in Eastern Canada are at greater distance, and typically can contribute only phase picks plus just a few correlations to the relocation of Charlevoix events. It is therefore easy to see why a criterion of "four or more" or "five or more" stations is hard to achieve for the events in this part of Eastern Canada: first, because greater epicentral distances are involved to reach four stations, and second because more than half the stations had to record the event with good signal-to-noise ratios and high similarity. For four out of the many more stations in the New Madrid region, good cross correlation is easier to achieve.

We conclude that the fraction of seismicity amenable to good relocations based on modern methods, ranges from about 10% across some broad regions to as much as 95% in other broad regions (major parts of Northern California). In some cases the percentage is influenced significantly by station density (and clearly this is the case for our study of China using a sparse network). We note that for purposes of monitoring a particular region—whether it be large such as North America or small such as a particular part of Liaoning Province, China—it is often true that archives of digital seismograms for events in that region are accumulating rapidly and will become more and more capable of supplying events that cross correlate. Modern methods are most suited to seismically active areas, since these allow cross-correlatable events to accumulate in a shorter time.

The choice of CC threshold has an influence on overall results, in that there is obviously a tradeoff between using as many relative arrival times measured from cross correlation as possible, but not using measurements of dubious quality. The choice of threshold is best made at a level that improves upon the accuracy of phase picks, and this resulted in thresholds of about 0.7 in our relocation studies using local stations. The DD algorithm eliminates outlier measurements, as long as they are not too numerous.

We note a conceptual inconsistency in that travel-time differences are determined by cross correlating with signals having a finite frequency band, and then measurements are passed to a location algorithm that assumes infinite-frequency ray theory. At event separations comparable to the shortest wavelength and comparable to the spatial scale of medium inhomogeneities, it may be more appropriate to use a location algorithm that explicitly recognizes the finite frequency of the underlying signals. A related point is that we have not explored the effects of material interfaces within the source region, and to the extent that such interfaces can substantially change the take-off directions of signals used for event location (as discussed for example in the case of fault-zone head waves by BEN-ZION and MALIN, 1991), there will be a distorting effect on relative locations, especially for events observed at a limited range of azimuths.

Although aspects of this work can now be regarded as quite routine, it continues to be of importance to acquire practical experience—for example, to provide feedback on questions such as: What station density is required to achieve enough CC measurements to locate the typical event; and what is the effect of archive longevity (how rapidly does the percentage of cross-correlatable events rise, as the data accumulate)? With such practical experience, it may then be appropriate to consider making modern methods of event location a part of standard operational practice, in support of publication of bulletins of seismicity which would then be enabled to attain considerable higher quality in seismically active areas where waveforms of neighboring events cross correlate.

The key to such an approach will be a commitment to long-term station operation, and building up waveform archives that can easily be used for large-scale searches. In practice, it is those network operators and their clients — who have already supported the effort to consistently collect, apply quality control, and build easily-usable archives for seismic waveforms — who are likely to benefit the most from the application of modern methods of event location.

Concerning computing power, we note that for a network consisting of tens of stations and thousands of events, a modern workstation is adequate. But for processing data in the amounts archived for example by the Northern California Earthquake Data Center (approximately a thousand stations, and a quarter of a million events), hardware with the necessary capability has only recently become widely available. The application of modern methods to major (terabyte scale) waveform archives, now being built for different networks and data centers, may require some further applied research at the scale of the full data set. A test bed that enables access to online archives could be the best facility to acquire such practical experience, prior to operational application of modern methods of seismic event location (waveform cross-correlation and multi-event location algorithms) on a large scale.

Once an archive has been established, and the signals of detected events have been associated and relocated by modern methods, new events can be accurately relocated relative to previously located events. After a period of time, say on the order of a year, all the events associated with the original waveform archive plus a year's worth of additional waveforms can be relocated all together to form a new archive, suitable as the reference set against which to locate the next year's events, and so on.

The context of this work is quite broad, since good locations are commonly the starting point for quantitative seismological studies including tomography and improved knowledge of Earth structure, seismic hazard analysis, earthquake physics and the interaction between neighboring events in a sequence, and the monitoring of explosions and earthquakes.

Finally we note that waveform cross correlation can be used not only to measure time differences and relative amplitudes very accurately, but to improve detection of small events, still allowing time differences to be measured even in cases where no phase picks can be made with confidence. See for example, GIBBONS and RINGDAL (2005). If waveform cross correlation can drop detection thresholds by about one magnitude unit, then approximately ten times more events may be detected, which in turn can lead (via multi-event location algorithms) to significant improvement in the location of all the events, including those of larger magnitude.

Acknowledgements

We greatly appreciate the assistance of network operators associated with the University of Memphis, the Geological Survey of Canada, the Northern California Seismic Network, and the Northern California Earthquake Data Center (NCEDC). In particular we thank David McCormack and Jim Lyons at the Geological Survey of Canada, Ottawa, who helped assemble phase and waveform data from earthquakes in the Charlevoix region; Jer-Ming Chiu of CERI, University of Memphis, and Jiakang Xie at Lamont-Doherty who helped us to acquire phase-pick and waveform data from the PANDA deployment at New Madrid. We also thank Mitch Withers at CERI, University of Memphis for New Madrid seismic network data and Doug Neuhauser for assistance with the NCEDC archive; and Yehuda Ben-Zion, Stephanie Ross and Malcolm Sambridge for constructive comments on the originally submitted manuscript. This work was supported by the Air Force Research Laboratory under contract F19628-03-C-0129. This is Lamont-Doherty Earth Observatory contribution number 6867.

References

Ben-Zion, Y., and Malin, P. (1991), *San Andreas fault zone head waves near Parkfield, California*, Science *251*, 1592–1594.

Bondár, I., Myers, S.C., Engdahl, E.R., and Bergman, E.A. (2004), *Epicenter accuracy based on seismic network criteria*, Geophys. J. Int. *156*, 483–496.

Chiu, J.M., Johnston, A.J., and Yang, Y.T. (1992), *Imaging the active faults of the central New Madrid seismic zone using PANDA array data*, Seismol. Res. Lett. *63*, 375–393.

Douglas, A. (1967), *Joint epicentre determination*, Nature *215*, 47–48.

Engdahl, E.R., van der Hilst, R., and Buland, R. (1998), *Global teleseismic earthquake relocation with improved travel times and procedures for depth determination*, Bull. Seismol. Soc. Am. *88*, 722–743.

Frohlich, C. (1979), *An efficient method for joint hypocenter determination for large groups of earthquakes*, Computers and Geosciences *5*, 3–4, 387–389.

Gibbons, S.J., and Ringdal, F. (2005), *The detection of rockbursts at the Barentsburg coal mine, Spitsbergen, using waveform correlation on SPITS array data*, NORSAR Semiannual Technical Summary, 1 July – 31 December, 2004, pp. 35–48.

Got, J.L., Fréchet, J., and Klein, F.W. (1994), *Deep fault geometry inferred from multiple relative relocation beneath the south flank of Kilauea*, J. Geophys. Res. *99*, 375–393.

Harris, D.B. (1991), *A waveform correlation method for identifying quarry explosions*, Bull. Seismol. Soc. Am. *81*, 2395–2418.

Hauksson, E., and Shearer, P. (2005), *Southern California hypocenter relocation with waveform cross correlation: Part 1. Results using the double-difference method*, Bull. Seismol. Soc. Am. *95*, 896–903.

Israelson, H. (1990), *Correlation of waveforms from closely spaced regional events*, Bull. Seismol. Soc. Am. *80*, 2177–2193.

Jordan, T.H., and Sverdrup, K.A. (1981), *Teleseismic location techniques and their application to earthquake clusters in the South-Central Pacific*, Bull. Seismol. Soc. Am. *71*, 1105–1130.

Peng, Z., and Ben-Zion, Y. (2005), *Spatiotemporal variations of crustal anisotropy from similar events in aftershocks of the 1999 M7.4 Izmit and M7.1 Duzce, Turkey, earthquake sequences*, Geophys. J. Int. *160*, 1027–1043.

Poupinet, G., Ellsworth, W.L., and Fréchet, J. (1984), M*onitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California*, J. Geophys. Res. *89*, 5719–5731.

Richards, P.G., Armbruster, J., Burlacu, V., Cormier, V.F., Fisk, M.D., Khalturin, V.I., Kim, W.-Y., Morozov, I.B., Morozova, E.A., Saikia, C.K., Schaff, D., Stroujkova, A., and Waldhauser, F. (2003), *Seismic Location Calibration for 30 International Monitoring System Stations in Eastern Asia: Final Results*, Proceedings, DOD/DOE Seismic Research Review, Tucson.

Richards-Dinger, K.B., and Shearer, P.M. (2000), *Earthquake locations in Southern California obtained using source-specific station terms*, J. Geophys. Res. *105*, 10,939–10,960.

Rowe, C.A., Aster, R.C., Phillips, W.S., Jones, R.H., Borchers, B., and Fehler, M.C. (2002), *Using automated, high-precision repicking to improve delineation of microseismic structures at the Soultz Geothermal Reservoir*, Pure Appl. Geophys. *159*, 563–596.

Rubin, A.M., Gillard, D., and Got, J.L. (1999), *Streaks of microearthquakes along creeping faults*, Nature *400*, 635–641.

Schaff, D.P., Bokelmann, G.H.R., Beroza, G.C., Waldhauser, F., and Ellsworth, W.L. (2002), *High-resolution image of Calaveras fault seismicity*, J. Geophys. Res. *107*, 2186, doi: 1029/2001JB000633.

Schaff, D.P., and Richards, P.G. (2004a), *Repeating seismic events in China*, Science *303*, 1176–1178.

Schaff, D.P., and Richards, P.G. (2004b), *Lg-wave cross correlation and double-difference location: Application to the 1999 Xiuyan, China, sequence*, Bull. Seismol. Soc. Am. *94*, 867–879.

Schaff, D.P., Bokelmann, G.H.R., Beroza, G.C., and Ellsworth, W.L. (2004), *Optimizing correlation techniques for improved earthquake location*, Bull. Seismol. Soc. Am. *94*, 705–721.

Schaff, D.P., and Waldhauser, F. (2005), *Waveform cross correlation based differential travel-time measurements at the Northern California Seismic Network*, Bull. Seismol. Soc. Am., *95*, 2446–2461.

Shearer, P. (1997), *Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence*, J. Geophys. Res. *102*, 8269–8283.

SHEARER, P., HAUKSSON, E., and LIN, G. (2005), *Southern California hypocenter relocation with waveform cross correlation: Part 2. Results using source-specific station terms and cluster analysis*, Bull. Seismol. Soc. Am. *95*, 904–915.

THURBER, C., TRABANT, C., HASLINGER, F., and HARTOG, R. (2001), *Nuclear explosion locations at the Balapan, Kazakhstan, nuclear test site: The effects of high-precision arrival times and three-dimensional structure*, Phys. Earth Plan. Int. *123*, 283–301.

WALDHAUSER, F., and ELLSWORTH, W.L. (2000), *A double-difference earthquake location algorithm: Method and application to the Northern Hayward fault, California*, Bull. Seismol. Soc. Am. *90*, 1353–1368.

WALDHAUSER, F. (2001), *HypoDD: A program to compute double-difference hypocenter locations*, U.S.G.S. Open-file Report, 01–113, Menlo Park, California.

WALDHAUSER, F., ELLSWORTH, W.L., and COLE, A. (1999), *Slip-parallel seismic lineations along the Northern Hayward fault, California*, Geophys. Res. Lett. *26*, 3525–3528.

WALDHAUSER, F., SCHAFF, D.P., RICHARDS, P.G., and KIM, W.-Y. (2004a), *Lop Nor revisited: Underground nuclear explosion locations, 1976–1996, from double-difference analysis of regional and teleseismic data*, Bull. Seismol. Soc. Am. *94*, 1879–1889.

WALDHAUSER, F., SCHAFF, D.P., KIM, W.-Y., and RICHARDS, P.G. (2004b), *Improved characterization of seismicity and fault structure by wide area event relocation*, EOS, Trans. Am. Geophys. U., 85, Fall Meet. Suppl., Abstract S53C–07.

WITHERS, M., ASTER, R., and YOUNG, C. (1999), *An automated local and regional seismic event detection and location system using waveform correlation*, Bull. Seismol. Soc. Am. *86*, 657–669.

XIE, J., LIU, Z., CONG, L., HERRMANN, R.B., and CHIU, J.M. (1997), *Rupture properties of clustered microearthquakes near intersecting intraplate faults of the New Madrid Seismic Zone: Implications on fault weakening*, J. Geophys. Res. *102*, 8,187–8,202.

XIE, J. (2001), *Rupture characteristics of clustered microearthquakes and variations in fault properties in the New Madrid Seismic Zone*, J. Geophys. Res. *106*, 26,495–26,509.

(Received May 26, 2005; accepted September 27, 2005)