

Lg-Wave Cross Correlation and Double-Difference Location: Application to the 1999 Xiuyan, China, Sequence

by David P. Schaff and Paul G. Richards

Abstract A surprising discovery has been made that in some cases the complex, highly scattered *Lg* wave is found to be similar for clusters of events. We analyze in detail a subset of 28 out of 90 events from the 1999 Xiuyan sequence. Cross correlations provide highly accurate differential travel-time measurements. Their error estimated from the internal consistency is about 7 msec. These travel-time differences are then inverted by the double-difference technique to obtain epicenter estimates that have location precision on the order of 150 m. The locations are computed with waveform data from four to five regional stations 500 to 1000 km away. The epicenter estimates are not substantially affected by the sparseness of stations or large azimuthal gaps. Comparison with a surface trace a few kilometers away and location estimates based on much more dense networks led us to conclude that the absolute positions are accurate to the 5-km level. 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 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).

Introduction

It is remarkable that after about three decades of experience with digital seismic data, the principal methods of event location are still based on traditional picking of the arrival times for specific seismic phases. In practice this means, at least for event location, that only a very small fraction of the available information contained in a modern seismogram is being used. Here, we evaluate the waveform cross-correlation method in application to event location on the regional scale. We choose for a case study the 1999 Xiuyan sequence in China, because it is a well known and much studied set of events and also because there is a need for additional ground truth events in China for calibration of stations in the International Monitoring System (IMS).

The use of phase picks for event location has been the subject of much recent research, emphasizing methods of application that reduce model error. For example, there has been significant reduction in location errors and in the size of confidence ellipses, when source-specific station corrections are applied to the interpretation of arrival times reported by stations of a sparse monitoring network as reported in the Annual Proceedings of IMS Location Calibration Workshops organized in Oslo by NORSAR for 1999 through 2003. But, although there is significant progress

here, it is understood that the uncertainty in event locations, expressed by the size of confidence ellipses, is determined by both model error and pick error. With pick errors for *Pn* arrivals being on the order of 1 sec, there is a basic limitation on the amount of error reduction achievable when model error alone is reduced; for example, by the double-difference technique, which reduces model error over large regions without the use of station corrections (Waldhauser and Ellsworth, 2000). Our goal in this article is to investigate the degree of location improvement on a regional scale that may be expected when pick error is greatly reduced in addition to model error. Our preferred method of reducing pick error is to use waveform cross correlation and of reducing model error, the double-difference algorithm.

Whether phase picking is automated or done by an analyst, there can be difficulties with phase identification, and phase picks are particularly poor for later-arriving phases. Thus, errors for *Sn* picks can exceed 20 sec, and the pick for *Lg* is often magnitude-dependent. But these late arrivals are often the largest phases, arriving over a window of several tens of seconds duration, and are therefore particularly well suited to use in waveform cross correlation. In practice waveform cross correlation can be used not only to measure

time differences 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.

Waveform cross-correlation methods have been applied previously by many investigators (for example, Poupinet *et al.*, 1984; Israelsson, 1990; Harris, 1991; Got *et al.*, 1994; Waldhauser *et al.*, 1999; Rubin *et al.*, 1999; Withers *et al.*, 1999; Thurber *et al.*, 2001; Schaff *et al.*, 2002; Rowe *et al.*, 2002; Fisk, 2002). Most of these studies focused on local earthquakes, and a few used regional and teleseismic waveforms to locate small clusters of events. In this and a companion article (Schaff and Richards, 2004), we explore the application of waveform cross correlation for improving location and detection on a broad regional scale. Our work is part of a general effort to improve the capability to locate seismic events accurately by using data from stations of the IMS. We have focused on China as a region of quite active but diffuse seismicity, for which a sparse network (the China Digital Seismograph Network) has been in operation since the mid-1980s. The techniques presented here should be of general research interest to those working in areas with a sparse regional network.

The 1999 Xiuyan sequence in Liaoning Province, China, shown in Figure 1, is especially interesting as a case

study for application of these methods, because it is the site of a successful short-term prediction of the 29 November M 5.9 mainshock (Jiao *et al.*, 2002). The prediction was issued on the basis of many precursory phenomena, including foreshocks, groundwater chemistry, geoelectricity, and tilt (Jiao *et al.*, 2002). It occurred in the general vicinity of the Haicheng earthquake of 4 February 1975, which was also successfully predicted on the basis of a series of foreshocks and changes in groundwater (Qidong *et al.*, 1981).

Data

We make use of data that are openly available at stations archived by the Incorporated Research Institutions for Seismology (IRIS). We also use the hypocentral information for the 90 events in the Xiuyan sequence reported by the Annual Bulletin of Chinese Earthquakes (ABCE) to select data windows and stations. The stations examined are BJT, ENH, HIA, MDJ, SSE, and XAN from the New China Digital Seismograph Network (CDSN) and INCN and ULN of the Global Seismograph Network (GSN-IRIS/USGS Albuquerque Seismological Laboratory). Stations ENH, SSE, and XAN were not used because of poor signal-to-noise ratios. The channels and components for which measurements were made were BHZ, BHN, and BHE with sample intervals of

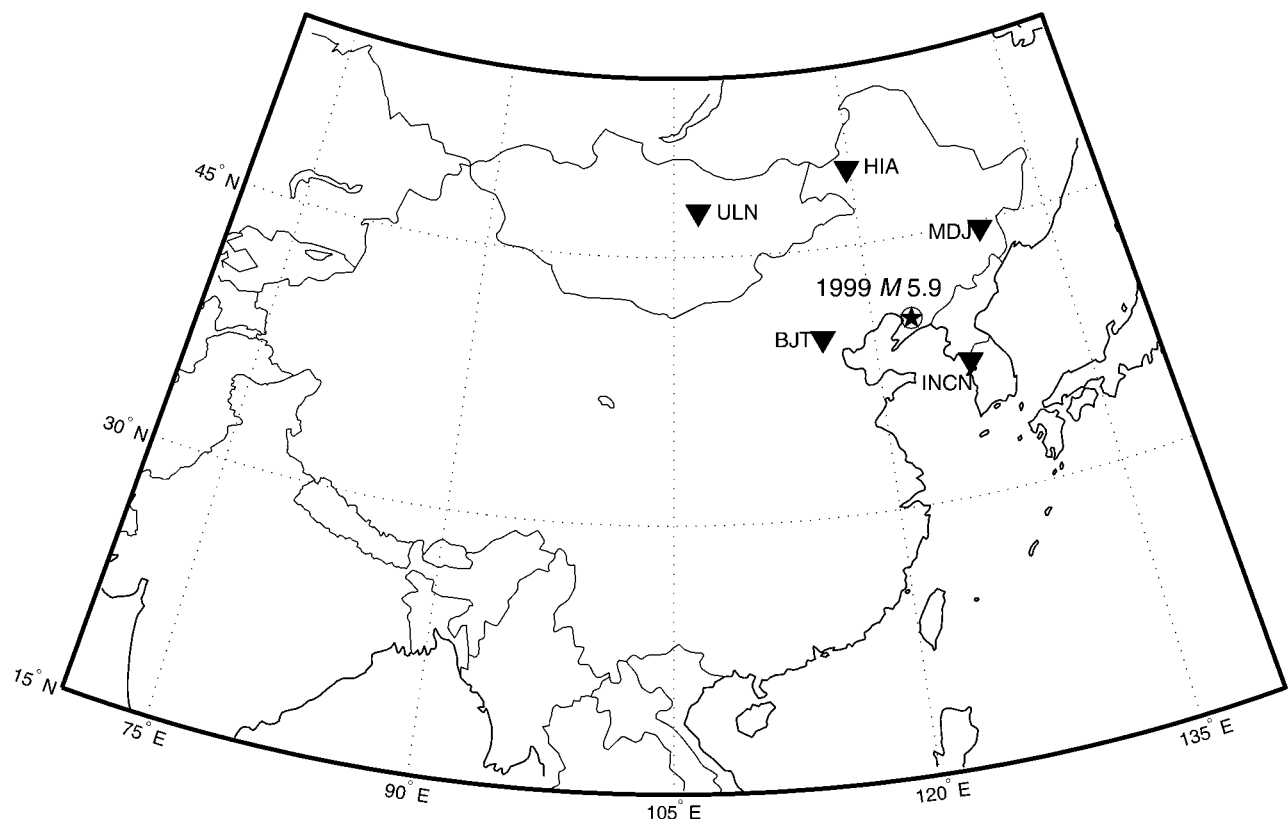


Figure 1. Star shows general location of 90 events in Xiuyan, China, known for a foreshock sequence enabling prediction of the 1999 M 5.9 mainshock. Stations shown recorded highly similar Lg waves and are used for the relocations.

0.05 sec. For the *Lg* correlations all the seismograms were filtered from 0.5 to 5 Hz to capture the dominant frequency band for this phase.

Technique

We improve on a recently developed highly efficient waveform cross-correlation algorithm for locally observed phases (Schaff *et al.*, 2004) to measure accurate travel-time differences for *Lg* phases observed at regional distances. The improvements mainly include interactive selection of the correlation windows and visual verification of quality in the resulting alignments. These data are then inverted by a double-difference technique (Waldhauser and Ellsworth, 2000; Waldhauser, 2001) to obtain precise relative locations for clusters of events. The double-difference method minimizes model errors across large areas without the use of station corrections. Particular attention is paid to the assessment of measurement errors and location uncertainties. By matching the resulting high-precision relative locations with surface information such as active faults, it is possible to turn many of these events into GT5 (ground truth absolute location accurate to 5 km) quality or better, as described for phase pick data by Waldhauser and Richards (unpublished data). In the case of China, we make use of a recently published U.S. Geological Survey digitized fault map that is based on mapped surface traces compiled by the Geological Publishing House (Beijing, 1984–1993).

Cross Correlation

Most of the success and experience with correlation methods has been realized on the local scale (e.g., Waldhauser *et al.*, 1999; Waldhauser and Ellsworth, 2002; Schaff *et al.*, 2002). At first we tried applying the same automatic program on regional *P* and *S* waves that we had used for local events (Schaff *et al.*, 2002) but met with limited progress. Two main problems with this approach were (1) the signal-to-noise ratio was too low for most of the events in the ABCE at which we were looking and (2) the *P* and *S* phases typically correlate only over short windows on the order of 1 to 2 sec. The challenge with this last point is that the correlation windows must be given some initial alignments, usually arrival-time picks or based on theoretical travel times. A cross-correlation function can typically recover offsets only up to half the window length (Schaff *et al.*, 2004). Therefore, the window has to be long enough to recover offsets large enough for the potential misalignments and still capture enough coherent energy. Because both pick and model errors on a regional scale for *P* and *S* waves exceed half the window lengths for windows on the order of 1 to 2 sec, it is hard to use waveform cross correlation for these phases. If there is a low signal-to-noise ratio, theoretical travel times may fail to get the initial alignments close enough for the correlation measurement to return a useful value. An interactive procedure can be used, if the signal-to-noise ratio is high enough for the events and the pick error

for the starting alignments can be reduced to these levels. A manual choice of window length also helps ensure the dominant energy is included in these instances. Another problem with shorter windows, especially if they are misaligned to start with, is that they are more prone to cycle skipping (the possibility of choosing a side lobe of the cross-correlation function instead of the true maximum).

Given these practical challenges in the context of relocating several clusters in China, our procedure gravitated toward a method that used phases with energy arriving over longer windows and larger signal-to-noise ratios for the small-magnitude events. The two basic hurdles listed here were overcome with the discovery that *Lg* waves were surprisingly good candidates and seemed to correlate well in cases with similar mechanisms and event separation distances less than 1 km (this article and Schaff and Richards, 2004). At high frequencies (0.5–5 Hz) and with long windows from 20 to 40 sec, initial misalignments are not so much a problem for the window length versus offset issue. Also, because many cycles are present at different amplitudes and frequencies, the chance for cycle skipping to introduce severe outlier delay measurements is essentially eliminated. Longer windows also have a greater probability of capturing the desired phase if the predicted travel times differ significantly from those of the real Earth. All these features lend themselves to greater ease of automating a cross-correlation procedure for *Lg* waves on large sets of events, as compared with regional *P* and *S* waves. To explore and develop the method, however, we used an interactive correlation program for a smaller set of events in the Xiuyan sequence at a few stations. This enabled us to ensure that the choice of window alignments and lengths accurately framed the dominant *Lg*-wave energy packet and also to verify that the resulting correlation alignments caused the waveforms to be shifted correctly and that the estimated errors were consistent.

Double-Difference Epicenter Equations

Because *Lg* waves travel as horizontally guided waves in the crust, we solve only for the epicenters. For simplicity, stations and event coordinates are converted to a cartesian system. Straight ray paths are assumed using a group velocity of 3.3 km/sec. Even though the *Lg* wave does not consist of a single ray path as typically used in location, we hypothesize that using the group velocity of the major energy packet will produce a good epicentral location for the centroid of moment release. The general double-difference equation from Waldhauser and Ellsworth (2000) is

$$\frac{\partial t_k^i}{\partial \mathbf{m}} \Delta \mathbf{m}^i - \frac{\partial t_k^j}{\partial \mathbf{m}} \Delta \mathbf{m}^j = dr_k^{ij},$$

where $\mathbf{m}$ is the vector of hypocentral model parameters, Δ indicates perturbations to the model parameters, t is travel time for the pair of events i and j observed at station k , and

dr is the residual between the observed and calculated differential travel times.

Solving for epicenter or hypocenter with a fixed depth leaves only the local map coordinates x and y along with the origin time τ :

$$\frac{\partial t_k^i}{\partial x} \Delta x^i + \frac{\partial t_k^i}{\partial y} \Delta y^i + \Delta \tau^i - \frac{\partial t_k^j}{\partial x} \Delta x^j - \frac{\partial t_k^j}{\partial y} \Delta y^j - \Delta \tau^j = dr_k^{ij}.$$

The straight ray path formulation allows the partial derivatives to be expressed analytically. We have $V = 3.3$ km/sec, x_s and y_s as the station coordinates, and D as the distance to the station.

$$t = \frac{D}{V}; D = \sqrt{(x - x_s)^2 + (y - y_s)^2}$$

$$\frac{\partial t}{\partial x} = \frac{x - x_s}{V \sqrt{(x - x_s)^2 + (y - y_s)^2}}$$

$$\frac{\partial t}{\partial y} = \frac{y - y_s}{V \sqrt{(x - x_s)^2 + (y - y_s)^2}}.$$

A simple system of the double-difference equations is then set up for the events in an iterative procedure to solve for updates to the model perturbations, until the solution does not change within some prescribed tolerance. We note that the double-difference technique is the optimal method for utilizing Lg cross-correlation data, because it inverts the differential travel-time measurements directly. All location programs that use absolute arrival times (even if adjusted by correlation measurements) would not produce the results presented here because the uncertainty in choosing appropriate reference arrival times can be mapped into scatter of the locations. By taking travel-time differences, no choice of absolute arrival times is needed.

Application to Xiuyan Data

Figure 1 displays the location of the source region for 90 events in Xiuyan, China, as well as the digital stations archived by IRIS with good correlation measurements. From a subset of 28 events, two clusters (one with 13 events and the other with 15 events) are observed that have good signal-to-noise ratios and similar Lg waves. The stations are from 500 to 1000 km away.

In Figure 2, the seismograms for six events in the first cluster display very similar Lg waves at station MDJ. The mean correlation coefficients (mcc) for each event are fairly high, almost all above 0.8. mcc is defined as the mean of the correlation coefficients for all event-pair measurements at that station associated with that particular event. From a correlation argument alone we can assume that these events are closely located, for example, by the quarter wavelength rule

(Geller and Mueller, 1980). Given the high frequencies and intense scattering over long windows for the Lg wave, it is likely that the events must occur close to each other with the same mechanism to share such similarity. The internal consistency from the redundancy of the delay measurements is 7 msec (σ_m), or about a tenth of a sample. This is obtained from the inversion of the set of overdetermined constraint equations at a station ($t_{ik} = t_{ij} + t_{jk}$, where $t_{ij} = t_j - t_i$ are the differential travel times measured by cross correlation, representing the difference in the individual event travel times to a common station, t_i). The standard deviation of the residual vector then gives an estimate of the measurement error (σ_m). This measurement precision is quite remarkable on a regional scale for this type of wave, because it is similar to the millisecond precision obtained for the best repeating events on a local scale (Poupinet *et al.*, 1984). Subsample precision is achieved by fitting a parabola to the peak of the cross-correlation function and by using the analytic maximum to compute the correlation coefficient and lag. The reason for this quality is that the measurement is quite robust due to the high similarity over the long windows.

To do the relocation, a simple system of double-difference equations is solved only for the epicenter and origin time, assuming straight ray paths and a velocity of 3.3 km/sec. The results in the right panel of Figure 3 indicate that the relative locations have precision much better than 1 km. The locations were calculated using only Lg -correlation data at four regional stations, BJT, HIA, MDJ, and ULN, which have an azimuthal gap of 220° . The starting locations for the iterative procedure were randomly perturbed several times by up to 100 km, converging to the same final locations, indicating they are quite robust and well constrained by the data. The standard deviation of the location residuals is about 0.02 sec which includes both measurement and model error and so is higher (as expected) than just σ_m alone. The major and minor axes of the 95% confidence ellipses are 100 m and 50 m, respectively. The bootstrap errors obtained by resampling the residual vector, shown by scattered small circles with different shades of gray for different events, agree well with the formal error ellipses. One comment is that, although all the waveforms in Figure 2 display a high degree of similarity, the top three (events 3, 8, and 4) show a greater number of features similar with each other. Likewise, greater internal similarity is also seen for the bottom three waveforms (events 6, 9, and 7). After relocation, it is seen that these event subclusters correspond to the three western-most events and the three eastern-most events, respectively (Fig. 3).

The left panel of Figure 3 compares the locations of these same events with locations obtained by applying the double-difference technique to traditional P -wave phase picks observed at a local/regional network of approximately 1000 stations. This latter analysis was done jointly by Lamont and Chinese seismologists (Chen Yun-tai and Yang Zhi-xian). Typically, each event was picked at several hun-

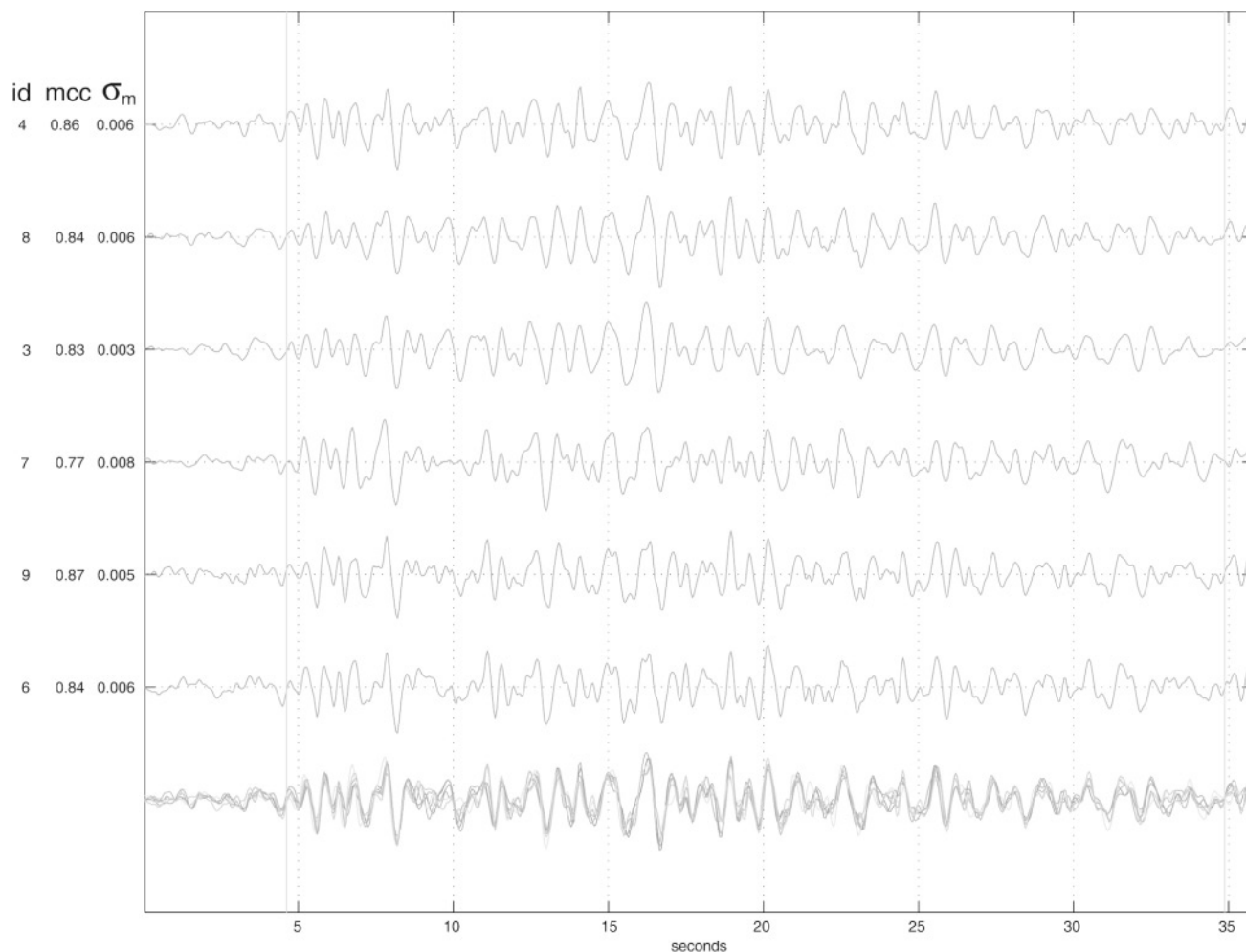


Figure 2. Highly similar *Lg* waveforms filtered from 0.5 to 5 Hz for a subset of the first cluster at station MDJ. Bottom trace shows the events superposed. Window length is about 30 sec denoted by vertical lines. The correlation coefficients are high (*mcc*) and the corresponding delay measurements have very good internal consistency (σ_m).

dred stations. The results are generally consistent but indicate considerable improvement using only the waveforms from the sparse network due to significantly increased measurement precision and the fact that *Lg* is such a slow wave. Although error estimates are not yet available for the events located with a local/regional network using only *P*-wave picks (left panel), the relative scatter seems to indicate that the resulting locations are good to at least the 1-km level, which we denote as RGT1 (relative ground truth). The locations and error estimates from the regional *Lg* cross-correlation measurements indicate a precision here of about 100 m or about an order of magnitude better than the *P*-wave phase data. Since both methods use the double-difference technique to minimize the effects of model error, most of the improvement is due to increased measurement precision. Because the absolute position of these events seems to match a fault trace on the surface to within a few kilometers, the locations are considered to be accurate at the 5-km or

better level (GT5) and are suitable ground truth events for calibration of IMS stations. Since the *Lg* wave is the largest amplitude phase on regional seismograms, this procedure may prove useful for improving locations of other small-magnitude events with only a few stations.

Locations for all the events in the first cluster, which includes the previously mentioned subset of six events, are shown in Figure 4. These are performed again by *Lg* correlation and double-difference epicentral location, except the similarity threshold is relaxed to allow more events to be relocated. We note that not all the events are recorded at all four stations. A mechanism for a *M* 5.1 aftershock in the Harvard centroid moment tensor (CMT) catalog (Dziewonski *et al.*, 1981) on this fault independently confirms the orientation of the strike that the epicenters suggest. These locations are not considered as precise as the previous subset because of the looser criteria and the joint inversion involving lower quality events and so some small shifts are ob-

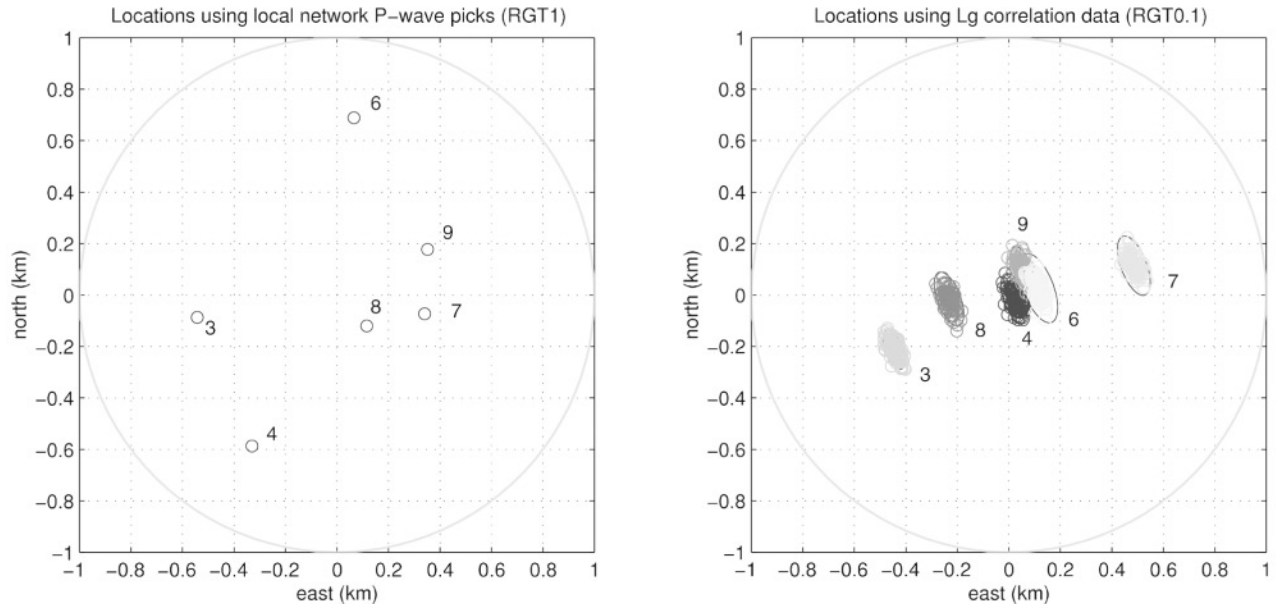


Figure 3. Comparison of double-difference relative locations for a subset of the first cluster for a local/regional network (left) using only P -wave phase picks recorded at several hundred stations and for a sparse regional network using Lg cross-correlation measurements (right). The rms travel-time residuals are about 1 sec for the P waves and 0.02 sec for Lg . The 95% confidence formal error ellipses and bootstrap errors (shaded small circles) are in good agreement (right). Origin in each case is taken as the centroid of the cluster.

served in the original set of six events. They are, however, useful for illuminating the fault trend on a larger scale and the location precision (which is consistent with Fig. 3) is still well below 1 km. The overall trend of the line of events shown in Figure 4 is independently corroborated by local P -wave double-difference locations by two separate groups (Chen Yun-tai and Yang Zhi-xian, and W. Chan, personal comm.), but since the scatter is closer to the kilometer level, their locations are not shown here for comparison.

Discussion

From the above-mentioned Lg correlation and double-difference locations it appears that the original hypothesis (namely assuming a single ray path and propagation velocity) is sufficient for giving high-precision epicentral locations of select events. Because the correlation measurements operate on a window of energy we infer that the appropriate group velocity is some average over the range of group velocities present in a ~ 30 -sec window surrounding the dominant packet of energy. Also, because the measurements are computed for a finite window length, we assume that the resulting locations represent the centroid of moment release as opposed to the actual epicenter associated with rupture nucleation. More research is warranted based on these results to better understand and model all the underlying causes and factors that would influence Lg correlation and locations at this scale.

Separation of Error Sources

Because the location residuals of 0.02 sec from Figure 3 are higher than 7 msec and combine both measurement and model error, it would be useful to understand all the potential sources for model error. Therefore, it may be good to also invert for the best-average group velocity instead of just choosing a single value of 3.3 km/sec. Note that group velocity variations along the ray paths to each station are canceled out by taking the double difference, and so the locations are sensitive only to group velocity perturbations at the source. If a single group velocity is used and the ray paths for each event in a doublet to a particular station are parallel, which they effectively are in this case, the result of changing the group velocity does not alter the overall structure of the seismicity but simply scales the picture by an isotropic expansion or contraction (Got *et al.*, 1994). Another possible influence of model error arises if the dominant Lg -wave packet does not travel exactly along a great circle path, in which case perturbations to the azimuth could also be inverted for and/or modeled.

With these single-phase measurements, there is always the possibility of station-clock errors. If present, it appears that station-clock errors are not a significant error source for these particular events. If they are, however, it may be possible to also use P - and S -wave correlation measurements now that similar event clusters have been identified. An initial look for these events reveals the signal-to-noise ratios are still quite low for these events and phases. If they can

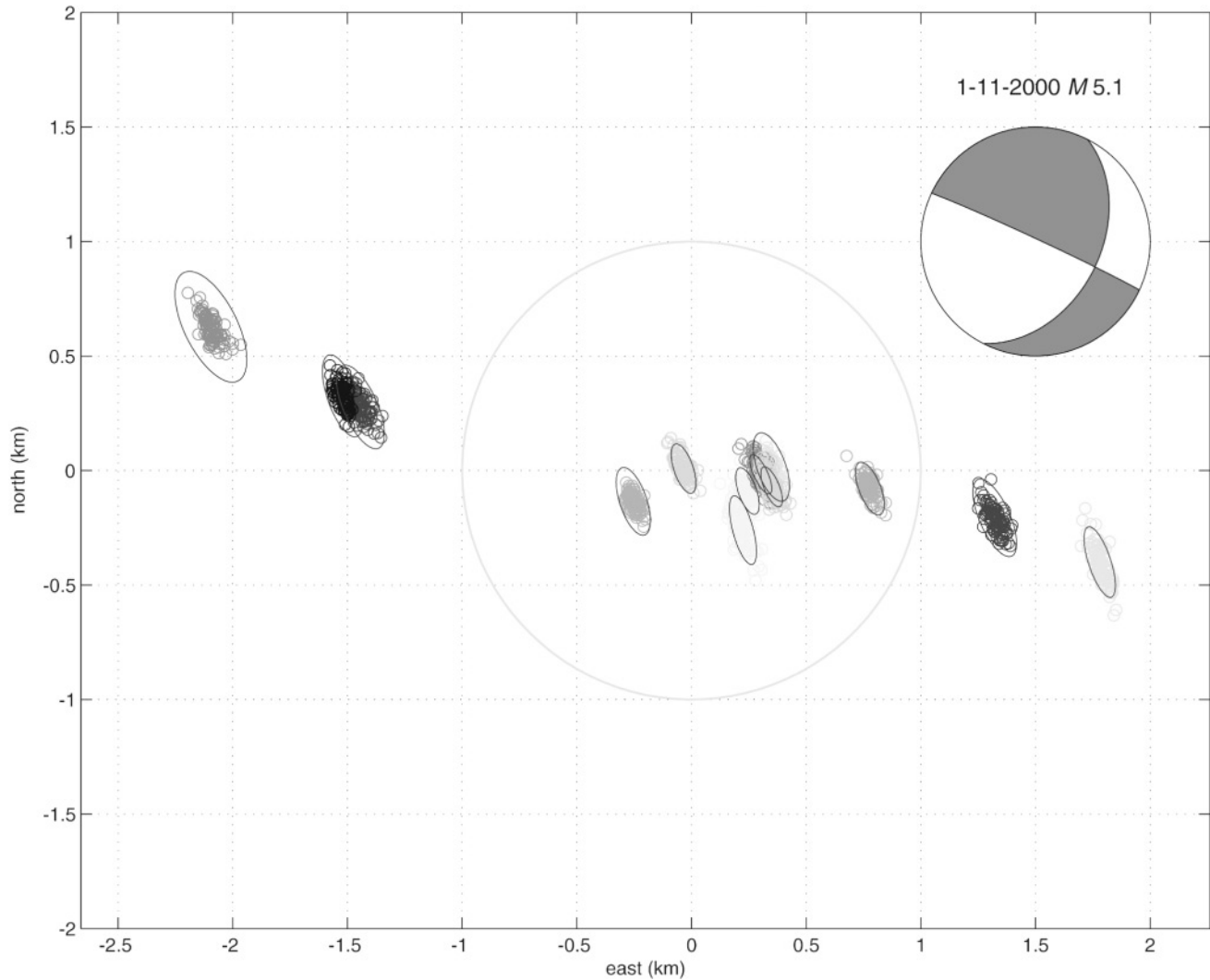


Figure 4. L_g correlation and double-difference relocation for all 13 events in the first cluster, which includes the subset of six events shown in Figures 2 and 3. The mechanism for a M 5.1 event from the Harvard CMT catalog (Dziewonski *et al.*, 1981) provides independent evidence on the strike of the fault.

be obtained, then use of $L_g - P$ and $L_g - S$ times would alleviate any errors due to station clocks.

Factors Controlling Waveform Similarity

For two seismograms to be similar they must share several common factors, including a common station response, similar path through the Earth, the same mechanisms, and source time functions. We are comparing seismograms at the same station for events that are separated at the 1-km level for station distances from 500 to 1000 km away. In the sequence of events at Xiuyan, we have observed two distinct clusters that correlate well within the cluster but have little similarity between the two clusters. The top panel of Figure 5 displays the cross-correlation coefficient (CC) matrix for these two clusters measured on the L_g waves recorded at station BJT. Clust1 is the set of 13 events that is shown in

Figure 4. Clust2 has 15 events that form a separate cluster. The maximum off-diagonal CC for the matrix is 0.94. A question is then raised as to why there are two distinct similarity clusters, if the events are occurring in the same general locale. Are different depths for the two sets the controlling factor for why the L_g similarity breaks down or is the breakdown due to different mechanisms?

To answer this question, we performed epicentral double-difference locations for the second cluster of 15 events as well. The L_g -correlation-derived locations are plotted in the lower panel of Figure 5. A list compiling the high-precision epicenters for both clusters totaling 28 events is given in Table 1. The absolute locations of the clusters are determined by using the centroid for each cluster as obtained from the local network locations based on 42 digital stations in Liaoning Province (Jiao *et al.*, 2002; W. Chan, personal

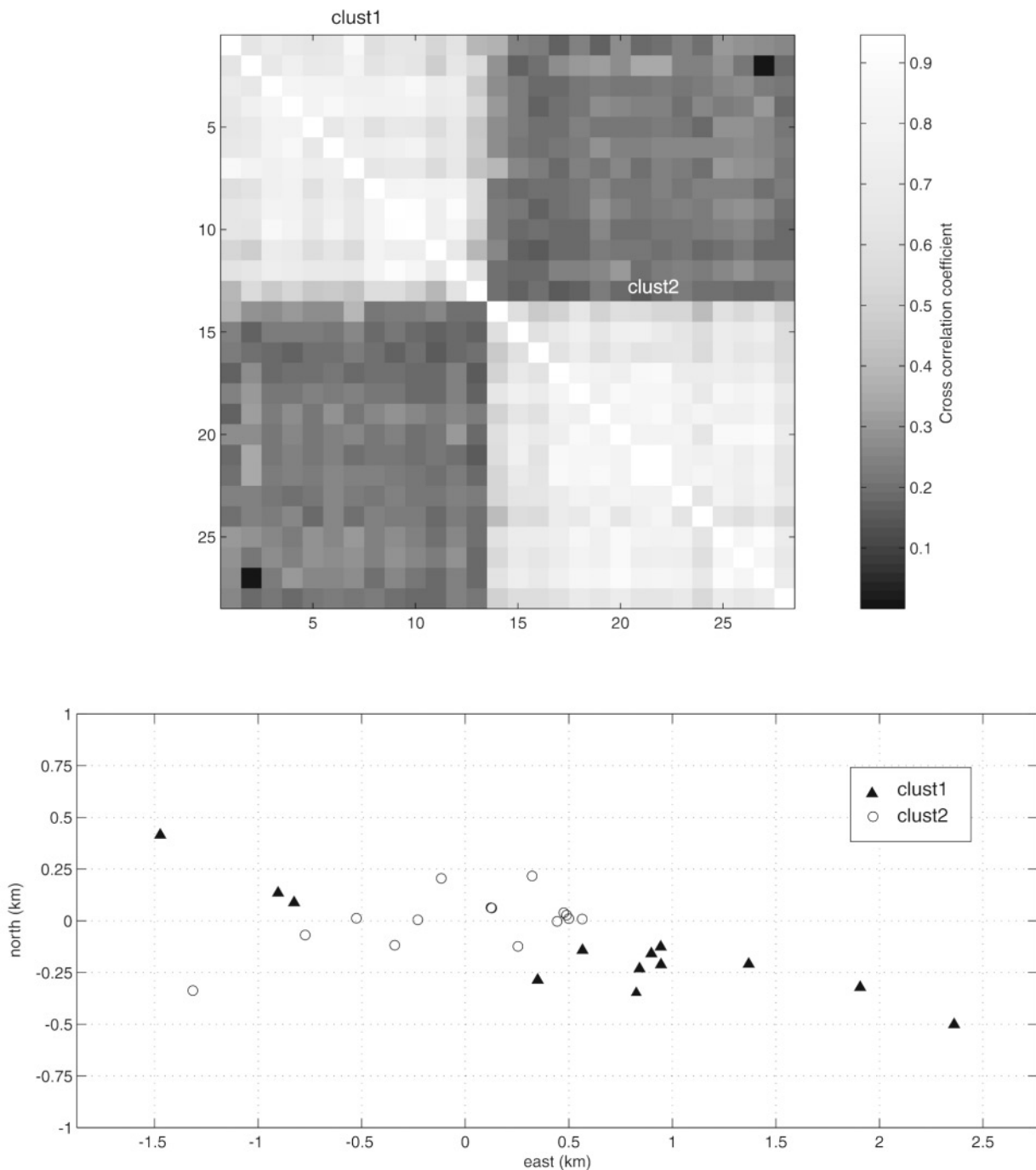


Figure 5. Of 90 events 28 were analyzed by cross correlation of the L_g wave. (top) Matrix of correlation coefficients for measurements made at station BJT showing two distinct similarity clusters. (bottom) Double-difference epicentral locations using the differential travel time data from the L_g correlations revealing two separate fault trends.

comm.). The plot shows that the two clusters are not separated by a substantial amount in map view; in fact, the two clusters are interweaved. Therefore, it does not appear that different paths through the Earth due to different locations is the explanation for the two clusters. What is suggested from the locations, however, is that the events from each cluster apparently lie on different fault planes indicative of

different mechanisms. A least-squares fit to the epicenters of each cluster suggests a difference in fault plane orientation of about 20° . Examination of the local network double-difference locations shows no discernible separation of the two clusters in depth (see Table 1), again arguing against different positions being the cause for the difference in waveforms. The L_g wave is a highly complex, scattered

wave that leaves the focal sphere from many different take-off angles and possibly azimuths (Anderson and Cormier, 1996), so we infer that even slight differences in mechanism will cause a breakdown in the similarity of *Lg* waves, even if the events are colocated and everything else is the same. Therefore, in the case of these Xiuyan events we conclude that the different similarity clusters actually reflect different mechanisms and give additional information as to on which fault plane the events at the intersection actually lie. In general, we expect depth to play a big role in the similarity of *Lg* waves, because they are a superposition of multiply reflected waves, but this does not appear to be the dominant factor in this case.

Detection Capabilities

The use of cross-correlation measurements is helpful not only in obtaining more accurate differential travel times, but it can also be used to improve detection capabilities. Detection of smaller-magnitude events is another important component of monitoring research. An exploration of *Lg* correlation and its utility for improving detection is developed more fully in Schaff and Richards (2004). Here, we

note that of the 28 events shown on Figure 5, only four are found in the Reviewed Event Bulletin (REB, now included within the ISC bulletin available at www.isc.ac.uk; see Table 1). For the 24 missing events we are able to find magnitudes in the local/regional data for 22 of them (C. Yun-tai, personal comm.). Table 2 lists the distribution of these missing events, the largest one being a *M* 4.9. Having access to only a sparse set of regional stations, it is possible to detect all these events based on the similarity of the *Lg* waves (Fig. 5, top panel). We note that if a larger event in the region is recorded and detected, it can be used to automatically cross correlate for smaller events that ordinarily would not be detected.

Sensitivity to Azimuthal Gap

The events in Figures 3, 4, and 5 were located with only four stations and a large azimuthal gap of 220°. Adding the Korean station INCN shown in Figure 1 reduces the gap to 133°, but unfortunately this station recorded none of the events in the first subcluster shown in Figure 3 and only eight events within the second cluster of Figure 5. A subset of four of these events met the more strict similarity require-

Table 1
High-Precision Epicentral Locations for 28 Events in the Xiuyan Sequence

ID	Date (mm/dd/yy)	Time	Latitude	Longitude	Depth* (km)	Mag*	$a^{\dagger}$ (m)	$b^{\dagger}$ (m)	$\theta^{\dagger}$ (°)	rms [‡] (sec)	Cluster	REB [§]
1	11/29/99	07:12:39.781	40.53536	123.01606	1.7	3.6	158	63	335	0.032	1	—
2	11/29/99	08:16:47.955	40.53813	123.00853	8.9	5.0	212	98	332	0.045	1	yes
3	11/30/99	06:06:55.171	40.53223	123.02980	9.7	4.9	125	50	341	0.033	1	—
4	11/30/99	06:09:35.266	40.53278	123.03555	7.2	4.3	86	33	340	0.021	1	—
5	11/30/99	17:47:02.971	40.53211	123.04800	11.1	4.2	147	56	335	0.033	1	—
6	11/30/99	13:07:05.678	40.53341	123.03622	6.9	3.7	76	26	334	0.016	1	—
7	11/30/99	12:19:45.717	40.53303	123.04172	13.6	4.3	99	40	338	0.025	1	—
8	12/01/99	04:45:31.041	40.53352	123.03233	11.8	4.3	91	36	341	0.023	1	—
9	12/02/99	15:16:21.997	40.53296	123.03677	10.5	3.6	75	28	337	0.017	1	—
10	12/03/99	09:57:52.502	40.53575	123.01516	9.9	3.7	151	57	340	0.035	1	—
11	12/12/99	21:49:31.114	40.53062	123.05329	9.8	4.1	129	41	341	0.022	1	—
12	01/09/00	11:41:36.616	40.53132	123.03532	1.5	3.7	123	37	344	0.018	1	—
13	01/12/00	05:00:35.592	40.53369	123.03676	10.5	4.3	127	53	340	0.034	1	—
14	11/27/99	07:48:02.886	40.53335	123.02736	9.1	3.2	116	18	336	0.010	2	—
15	11/29/99	05:05:13.829	40.53124	123.00893	11.3	3.4	85	19	335	0.012	2	—
16	11/29/99	06:55:45.638	40.53335	123.02034	8.7	3.3	54	14	336	0.009	2	—
17	11/29/99	16:29:19.045	40.53503	123.02581	9.8	3.2	36	12	335	0.008	2	—
18	11/29/99	14:29:09.452	40.53463	123.03023	7.9	3.2	31	10	336	0.006	2	—
19	11/30/99	20:33:00.449	40.53446	123.01817	10.0	4.3	42	15	334	0.010	2	—
20	11/30/99	20:58:58.981	40.53479	123.03008	11.5	3.5	35	12	335	0.008	2	—
21	12/21/99	03:34:31.384	40.53456	123.03102	11.4	3.2	42	19	337	0.011	2	—
22	12/27/99	11:57:25.640	40.53503	123.02584	11.8	3.2	48	19	340	0.012	2	—
23	12/27/99	15:48:37.480	40.53450	123.02161	11.7	3.7	47	25	338	0.017	2	yes
24	12/27/99	11:27:12.392	40.53378	123.01523	8.9	4.0	53	22	336	0.015	2	yes
25	02/01/00	16:07:20.118	40.53625	123.02297	9.9	3.6	58	20	334	0.012	2	—
26	03/19/00	15:41:53.560	40.53444	123.02958	—	—	51	27	338	0.018	2	yes
27	03/19/00	15:18:26.984	40.53482	123.02995	—	—	29	15	338	0.010	2	—
28	03/19/00	17:47:03.562	40.53643	123.02813	—	—	79	10	337	0.006	2	—

*From double-difference catalog of Chen Yun-tai and Yang Zhi-xian (personal comm.).

[†]Parameters for 95% confidence error ellipses (a is semimajor axis, b is semiminor axis, and θ is angle to a with respect to north).

[‡]rms differential travel-time residual.

[§]Event present in Reviewed Event Bulletin.

Table 2
Missing Events in the REB

Magnitude	No.
3 to 3.5	7
3.5 to 4	7
4 to 4.5	7
4.5 to 5	1

ments imposed for Figure 3. The double-difference locations for these events are shown in Figure 6 with error ellipses that are comparable in dimension to those obtained earlier. Note the location precision and observed structure are well below the 1-km level, as in Figure 3. The measurement error is 6 ms, whereas the location residuals are 0.025 sec. In this case, four stations were used, corresponding to the smaller azimuthal gap of 133° (BJT, HIA, INCN, and MDJ). To test the sensitivity of these locations to a greater azimuthal gap of 220° , we leave out the station INCN and perform the locations again which are plotted as small stars. Relative position vectors are also shown between the two sets of locations with corresponding lengths in meters from west to east: 117, 84, 147, and 62. The shifts are on the order of the error bars and are all less than 150 m. From this analysis it appears that the effect of sparse networks and large azimuthal gaps does not substantially contribute to mislocation of more than a few hundred meters by our *Lg* correlation and location method. The reason for this is that the double-difference technique is less sensitive to large azimuthal gaps, because it differences out large velocity variations between the sources and receivers. Requiring the data to be paired ensures that whatever biases are present are common and thus canceled out. The technique is sensitive only to variations in slowness (velocity and azimuth) over the source

area, which for this example are small (on the order of 1 km) compared with the station distance. As indicated previously, the greatest source of unmodeled error here probably comes from minor perturbations to the azimuth for paths departing the source slightly off the great circle path.

Advantages of *Lg*-Wave Correlation and Location

Lg-wave cross-correlation measurements, when available, seem to have many advantages for constraining epicenters compared with conventional methods. The first advantage arises when the waves are similar over windows of substantial length (10 or more sec). The measurement error (σ_m) can then achieve subsample precision and approaches the millisecond levels attained on a local scale for *P* and *S* phases. The second advantage is that the *Lg* wave is much slower than the *P* wave and thus constrains the locations better given the same error in measured travel time. A third advantage is that epicentral locations are controlled by the apparent horizontal velocity, which for the case of *Lg* is the same as the true wave speed, but for *P* and *S* waves it is even higher, in practice, than the body-wave velocity. A fourth advantage is that the *Lg*-wave is commonly the largest amplitude signal, and thus has the best signal-to-noise ratio.

We have shown that the pairwise measurements for two sets of similar *Lg* events studied show a high degree of internal consistency. From the redundancy of these measurements and a subsequent inversion for absolute arrival-time adjustments, the measurement error can be estimated to be on the order of 5 to 10 msec. The fact that these measurements are made on such long windows, including frequencies up to 5 Hz, makes them insensitive to cycle skipping and quite robust. These features contribute to the ease of automating the procedure that is discussed further in Schaff and Richards (2004). Also presented in Schaff and Richards

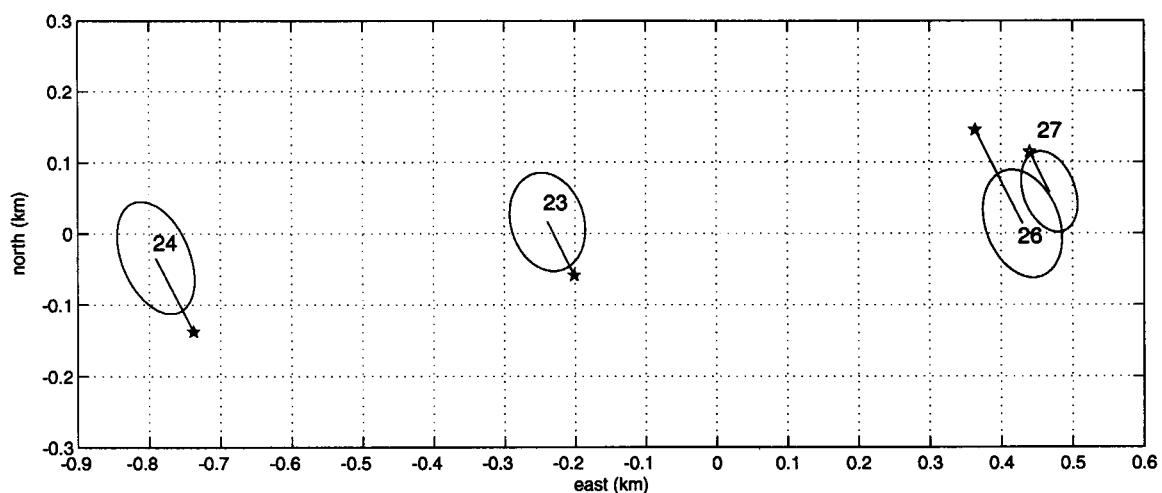


Figure 6. Double-difference epicentral locations for a subset of four events out of the second cluster using a station set with a smaller azimuthal gap of 133° (ellipses) compared with locations obtained previously for a larger gap of 220° (stars). Shifts are all less than 150 m.

(2004) is an empirical estimate of typical pick error found in the ABCE for various phases. For *Lg* arrivals, this measurement error is on the order of 10 to 20 sec, which is not surprising given the complex series of arrivals that are observed. These errors may be expected when trying to pick first breaks of such a wave in the coda of a prior *S* wave. These errors are also not unusual, even if the peak of the envelope for the *Lg*-energy packet is chosen, because of the increasing energy that comes in at later times and differences in envelope shape. For these reasons, *Lg* arrivals are often not even routinely picked.

The advantage of cross-correlation measurements for *Lg* (and other phases of interest) is that they are not restricted to just first arrivals but use a more substantial fraction of the information contained in the seismogram, including most of the radiated seismic energy (Shi *et al.*, 2000). As shown previously, these measurements can make a thousandfold improvement in measurement error alone for selected events. Our approach applies to about a third of the seismicity considered here for the Xiuyan sequence (28 of 90 events in the ABCE). To show it is not just an isolated instance, we have analyzed the possibility of *Lg* correlation on a large scale for about 14,000 events in the ABCE and show that with more restrictive criteria they correlate above a 0.8 threshold for at least 1301 (9%) of the events (Schaff and Richards, 2004). The usefulness of *Lg*-correlation measurements becomes even more pronounced for small-magnitude events that may not have a good enough signal-to-noise ratio to pick *P* and *S* arrivals.

For constraining epicenters, it is the horizontal (or apparent) velocity that governs the location through the simple equation, $D = V_{\text{app}} \times t$, where V_{app} now applies in the general case for *P*, *S*, and *Lg* waves. The measurement errors map linearly into location errors through $\sigma_D = V_{\text{app}} \sigma_m$. As noted earlier, given the same measurement error, the location errors are smaller if the wave speed is smaller. For *Lg* waves the horizontal velocity equals the true velocity, but for *P* and *S* waves, the apparent velocity is higher according to $V_{\text{app}} = V / \sin \theta$, where θ is the takeoff angle. Using $V = 6$ km/sec and $\theta = 45^\circ$ gives $V_{\text{app}} \cong 8.5$ km/sec. Therefore, using a slower group velocity for the *Lg* wave of 3.3 km/sec, the epicentral errors improve by a factor of 2.5. As the takeoff angles become steeper for teleseismic *P*, the reduction in errors is even more significant.

Conclusions

A surprising discovery has been made that the complex, multiply scattered *Lg* wave observed at regional distances can correlate extremely well over long windows at high frequencies in select cases. The resulting measurement error is similar to that achieved on a local scale for correlation of *P* and *S* waves. Table 3 provides a summary of measurement precision from our current understanding and experience on local and regional scales. The similarity of the waveforms is controlled by the same factors as for local seismograms, including common station, mechanism, and location (same paths through the Earth). For *Lg* waves, it appears that interevent separation distances up to a kilometer can provide useful cross-correlation measurements in the 0.5- to 5-Hz band. Since correlation measurements are known to break down with increasing separation distance (e.g., Menke *et al.*, 1990; Nadeau *et al.*, 1995; Schaff *et al.*, 2004), the measurement error (σ_m) in Table 3 is estimated for the ranges in column two starting at zero separation. In the study presented, it appears that mechanism is a more dominant factor in determining *Lg* similarity for these events than depth is. Based on preliminary results in China, it appears that *Lg* waves will provide useful correlation measurements for about 9% to 30% of the events in the ABCE.

The correlation measurements provide highly accurate differential travel-time data for *Lg* waves that then can be solved for epicenters using the double-difference technique, assuming straight ray paths and a group velocity of 3.3 km/sec. With this approach, it is demonstrated that relative epicenters can be determined with precision on the order of 150 m or better. This result is achieved even for a sparse regional network with large azimuthal gaps, which often arises in monitoring smaller-magnitude events. Comparison with local/regional network locations from collaborators in China, using more than 100 *P*-wave phase picks for each event (Fig. 3), although with more scatter (on the order of 500 m to 1 km), independently support our results. Since both sets of locations use the double-difference technique that reduces model error, the differences can be attributed primarily to differences in measurement error. The *Lg*-correlation measurements have extremely high measurement precision of 7 msec compared with typical *P*-wave pick error ranging from 0.1 to 1 sec. The other reason that the *Lg* data show

Table 3
Measurement Precision for Different Correlation Measurements

Phase	Separation Distance (km)	Window Length (sec)	Filter Bands (Hz)	σ_m (sec)
Local <i>P</i> and <i>S</i> waves	0 to 2	2	1 to 10	0.001 to 0.01
Regional <i>Lg</i> wave	0 to 1	20 to 40	0.5 to 5	0.001 to 0.01

an improvement over the *P*-wave data is that *Lg* has a much slower horizontal velocity, which provides better constraint for the epicenter.

Conventionally, ground truth events are designated only in terms of epicenter. Another sizeable component of discrimination is to improve depth estimates for events to sift out earthquakes from manmade explosions. *Lg* correlations and locations can not provide any constraints on the depth of the events, except for the fact that the mere presence of the *Lg* wave is usually an indication of a shallow focus. To constrain depth better, *P* and *S* waves should be examined along with various depth phases associated with modeling work (e.g., Dziewonski *et al.*, 1981; Wallace and Helmberger, 1982; Goldstein and Dodge, 1999; Du *et al.*, 2003). We appreciate that the high-resolution locations obtained here are relative and additional work is needed to convert these events to absolute locations, for example, by matching to surface fault maps.

These methods present significant potential advances in the field of monitoring research in better determination of relative epicentral location. Studies using *Lg* waves are especially important for small-magnitude events that lack a well recorded *P* or *S* wave. Due to the high frequencies and long windows over which *Lg* is found to correlate, the measurements are quite robust, which facilitates automation. When available, *Lg*-correlation data can offer up to a thousandfold increase in measurement precision as compared with traditional phase picks. As a result, the double-difference locations using only a handful of regional stations can surpass the location precision of denser local networks that use only *P*-wave phase data.

Acknowledgments

We thank our Chinese collaborators, Chen Yun-tai and Yang Zhixian, for providing double-difference locations using *P*-wave picks for comparison. We also thank Winston Chan for supplying similar double-difference locations for the Xiuyan sequence. We are grateful for helpful comments from Michael Antolik, Jose Pujol, and an anonymous reviewer. Our work was supported by Contract No. DTRA01-00-C-0031 of the Defense Threat Reduction Agency. This article is Lamont-Doherty Earth Observatory Contribution Number LDEO 6572.

References

- Anderson, T. S., and V. F. Cormier (1996). Regional variations in *Lg* observed and synthesized at the CNET and KNET arrays, *Seism. Res. Lett.* **67**, 29.
- Du, W.-X., W.-Y. Kim, and L. R. Sykes (2003). Earthquake source parameters and state of stress for the Northeastern United States and Southeastern Canada from analysis of regional seismograms, *Bull. Seism. Soc. Am.* **93**, 1633–1648.
- Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.* **86**, 2825–2852.
- Fisk, M. D. (2002). Accurate locations of nuclear explosions at the Lop Nor Test Site using alignment of seismograms and IKONOS satellite imagery, *Bull. Seism. Soc. Am.* **92**, 2911–2925.
- Geller, R. J., and C. S. Mueller (1980). Four similar earthquakes in Central California, *Geophys. Res. Lett.* **7**, 821–824.
- Goldstein, P., and D. Dodge (1999). Fast and accurate depth and source mechanism estimation using *P*-waveform modeling: a tool for special event analysis, event screening, and regional calibration, *Geophys. Res. Lett.* **26**, 2569–2572.
- Got, J.-L., J. Fréchet, and F. W. Klein (1994). Deep fault plane geometry inferred from multiple relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.* **99**, 15,375–15,386.
- Harris, D. B. (1991). A waveform correlation method for identifying quarry explosions, *Bull. Seism. Soc. Am.* **81**, 2395–2418.
- Israelsson, H. (1990). Correlation of waveforms from closely spaced regional events, *Bull. Seism. Soc. Am.* **80**, 2177–2193.
- Jiao, W., H. Gu, and G. Gu (2002). Review of the earthquake prediction for the 29 November 1999 M 5.9 Xiuyan earthquake by relocating the sequence with the double-difference method, *Seism. Res. Lett.* **73**, 259.
- Menke, W., A. Lerner-Lam, B. Dubendorff, and J. Pacheco (1990). Polarization and coherence of 5 to 30 Hz seismic wave fields at a hard-rock site and their relevance to velocity heterogeneities in the crust, *Bull. Seism. Soc. Am.* **80**, 430–449.
- Nadeau, R. M., W. Foxall, and T. V. McEvilly (1995). Clustering and Periodic Recurrence of Microearthquakes on the San Andreas Fault at Parkfield, California, *Science* **267**, 503–507.
- Poupinet, G., W. L. Ellsworth, and J. Fréchet (1984). Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California, *J. Geophys. Res.* **89**, 5719–5731.
- Qidong, D., J. Pu, L. M. Jones, and P. Molnar (1981). A preliminary analysis of reported changes in groundwater and anomalous animal behavior before the 4 February 1975 Haicheng earthquake, in *Earthquake Prediction an International Review*, D. W. Simpson and P. G. Richards (Editors), Maurice Ewing Series 4, American Geophysical Union, Washington, D.C.
- Rowe, C. A., R. C. Aster, W. S. Phillips, R. H. Jones, B. Borchers, and M. C. Fehler (2002). Using automated, high-precision repicking to improve delineation of microseismic structures at the Soutz Geothermal Reservoir, *Pure Appl. Geophys.* **159**, 563–596.
- Rubin, A. M., D. Gillard, and J.-L. Got (1999). Streaks of microearthquakes along creeping faults, *Nature* **400**, 635–641.
- Schaff, D. P., G. H. R. Bokelmann, G. C. Beroza, F. Waldhauser, and W. L. Ellsworth (2002). High-resolution image of Calaveras Fault seismicity, *J. Geophys. Res.* **107**, 2186, doi 10.1029/2001JB000633.
- Schaff, D. P., G. H. R. Bokelmann, G. C. Beroza, and W. L. Ellsworth. Optimizing correlation techniques for improved earthquake location, *Bull. Seismol. Soc. Am.*, in press.
- Schaff, D. P., and P. G. Richards (2004). Repeating seismic events in China, *Science* **303**, 1176–1178.
- Shi, J., P. G. Richards, and W.-Y. Kim (2000). Determination of seismic energy from *Lg* waves, *Bull. Seismol. Soc. Am.* **90**, 483–493.
- Thurber, C., C. Trabant, F. Haslinger, and R. Hartog (2001). Nuclear explosion locations at the Balapan, Kazakhstan, nuclear test site: the effects of high-precision arrival times and three-dimensional structure, *Phys. Earth Planet. Interiors* **123**, 283–301.
- Wallace, T. C., and D. V. Helmberger (1982). Determining source parameters of moderate-size earthquakes from regional waveforms, *Phys. Earth Planet. Interiors* **30**, 185–196.
- Waldhauser, F. (2001). HypoDD: a program to compute double-difference hypocenter locations, *U.S. Geol. Surv. Open-File Rept.* **01–113**, 1–25.
- Waldhauser, F., W. L. Ellsworth, and A. Cole (1999). Slip-parallel seismic lineations along the northern Hayward fault, California, *Geophys. Res. Lett.* **26**, 3525–3528.
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California, *Bull. Seism. Soc. Am.* **90**, 1353–1368.

- Waldhauser, F., and W. L. Ellsworth (2002). Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations, *J. Geophys. Res.* **107**, 2054, doi 10.1029/2000JB000084.
- Withers, M., R. Aster, and C. Young (1999). An automated local and regional seismic event detection and location system using waveform correlation, *Bull. Seism. Soc. Am.* **89**, 657–669.

Lamont-Doherty Earth Observatory of Columbia University
Palisades, New York 10964
dschaff@ldeo.columbia.edu
(D.P.S.)

Lamont-Doherty Earth Observatory of Columbia University
Department of Earth and Environmental Sciences
Palisades, New York 10964
richards@ldeo.columbia.edu
(P.G.R.)

Manuscript received 1 July 2003.