

# Optimizing Correlation Techniques for Improved Earthquake Location

by David P. Schaff,\* Götz H. R. Bokelmann, William L. Ellsworth, Eva Zanzierkia, Felix Waldhauser,\* and Gregory C. Beroza

**Abstract** Earthquake location using relative arrival time measurements can lead to dramatically reduced location errors and a view of fault-zone processes with unprecedented detail. There are two principal reasons why this approach reduces location errors. The first is that the use of differenced arrival times to solve for the vector separation of earthquakes removes from the earthquake location problem much of the error due to unmodeled velocity structure. The second reason, on which we focus in this article, is that waveform cross correlation can substantially reduce measurement error. While cross correlation has long been used to determine relative arrival times with subsample precision, we extend correlation measurements to less similar waveforms, and we introduce a general quantitative means to assess when correlation data provide an improvement over catalog phase picks. We apply the technique to local earthquake data from the Calaveras Fault in northern California. Tests for an example streak of 243 earthquakes demonstrate that relative arrival times with normalized cross correlation coefficients as low as  $\sim 70\%$ , interevent separation distances as large as to 2 km, and magnitudes up to 3.5 as recorded on the Northern California Seismic Network are more precise than relative arrival times determined from catalog phase data. Also discussed are improvements made to the correlation technique itself. We find that for large time offsets, our implementation of time-domain cross correlation is often more robust and that it recovers more observations than the cross spectral approach. Longer time windows give better results than shorter ones. Finally, we explain how thresholds and empirical weighting functions may be derived to optimize the location procedure for any given region of interest, taking advantage of the respective strengths of diverse correlation and catalog phase data on different length scales.

## Introduction

Two principal sources of error for any earthquake location problem are arrival time measurement error and errors arising from unmodeled velocity structure, subsequently referred to as model error. Progress in reducing both these error sources, most recently using the combined strengths of waveform cross correlation measurements and the double-difference location approach (Waldhauser and Ellsworth, 2000), has been realized on the Calaveras, Hayward, and San Andreas Faults with a reduction of 1–2 orders of magnitude in location error at small length scales (Rubin *et al.*, 1999; Waldhauser *et al.*, 1999; Schaff *et al.*, 2002). In these applications, initial earthquake locations from manual phase picks have all been dramatically improved using cross correlation measurements. In this article, we present the methods we used to enable more high-quality correlation mea-

surements to be utilized for the improvements in the case study on the Calaveras Fault (Schaff *et al.*, 2002). The general methodology should apply to other regions where correlation measurements are used to improve earthquake location.

The utility of waveform cross correlation together with relative location techniques is well established (Poupinet *et al.*, 1984; Fréchet, 1985; Ito, 1985; Frémont and Malone, 1987; Deichmann and Garcia-Fernandez, 1992; Got *et al.*, 1994; Dodge *et al.*, 1995; Shearer, 1997), but it is only recently that relative location techniques have been applied to large numbers of events at spatial scales of more than a few kilometers (see Table 1) (e.g., Rubin *et al.*, 1999; Waldhauser *et al.*, 1999; Rowe *et al.*, 2002; Schaff *et al.*, 2002). The reason for this is in part historical. Earlier work concentrated on small clusters of earthquakes such as foreshocks (Dodge *et al.*, 1996) or repeating events (e.g., Poupinet *et al.*, 1984) that were certain to have extremely similar waveforms due to separations on the order of 100 m or less. Recently, large

\*Present address: Lamont–Doherty Earth Observatory, Columbia University, Palisades, NY 10964.

Table 1  
Examples of Progress for Correlation-Based Relative Relocation over the Years

Group	Year	Region	Events Relocated
Poupinet <i>et al.</i>	1984	Doublets in Alps and Calaveras Fault	~2
Got <i>et al.</i>	1994	South flank fault on Big Island, Hawaii	~200
Rubin <i>et al.</i>	1999	Creeping section of San Andreas Fault	~3,200 (75%)
USGS/Stanford	2000	Hayward, Calaveras, San Andreas, Landers	~20,000 (90%)
USGS/Stanford/Lamont	2002–present	Entire northern California catalog	~250,000 (99%)

volumes of waveform data have become readily accessible at earthquake data centers, facilitating the processing of larger and more diverse waveform data sets (Neuhauser *et al.*, 1994).

An important step in applying waveform cross correlation to earthquakes at larger spatial scales is the recognition that correlation data can be a significant improvement over catalog data under more general circumstances than was previously appreciated. In this study, we use the postfit relative arrival time residuals from the double-difference location algorithm to assess independently the errors in the two data sets. From this quantitative comparison, we find that cross correlation measurements can be used with great success to much lower similarity thresholds than have been previously applied. For seismicity recorded by the Northern California Seismic Network (NCSN) instrumentation and telemetry, we find that, on average, cross correlation measurements provide more precise measures of relative arrival time differences than catalog picks at separation distances of up to 2 km.

At larger spatial scales, arrival time picks in the catalog provide essential information, and without them it would be impossible to relocate all of the seismicity. Thus, combining the two data sets is important. By using postfit residuals from the location algorithm to develop a rational data weighting for both the catalog and cross correlation arrival time measurements, we obtain locations that take advantage of the strengths of both data sets.

### Time-Domain Cross Correlation versus Cross Spectral Approach

Typical standard errors for catalog hypocentral locations in northern California range from hundreds of meters to a kilometer horizontally and about twice that vertically, which can lead to location scatter that obscures important details of fault behavior. The largest source of error arises from unmodeled earth structure, but errors in individual arrival time measurements also contribute significantly to the problem. In most cases *S*-wave picks are not made, and thus the constraint they provide on hypocentral locations, particularly in depth, is not available. Arrival times based on correlations, however, only need to be centered on a window of energy, even if the first breaks of the *S*-waves are not discernible and therefore not picked.

Arrival time data for a set of events considered simul-

taneously (the “catalog”) can better constrain the relative locations between events. In the hypoDD algorithm of Waldhauser and Ellsworth (2000), this was done by converting the arrival times into time differences between common phases of different earthquakes recorded at the same station. Because the phases traverse similar paths if two earthquakes are close to one another, their travel-time difference will not be affected by unmodeled structure for the parts of the propagation path that are common to the two events. Those travel-time differences can therefore be used to obtain precise relative location estimates by reducing velocity model error. To reduce the measurement error of the time differences, time-domain cross correlation or the cross spectral technique may be employed. The principles of delay estimation using these two methods are illustrated in Figure 1 (see also Jenkins and Watts [1968]). Time-domain cross correlation is defined as

$$c(\tau) = n \int u_1(t)u_2(t + \tau)dt \quad (1)$$

for seismograms  $u_1(t)$  and  $u_2(t)$  and with the normalization  $n = 1/[\int u_1^2(t)dt \int u_2^2(t)dt]^{1/2}$ . The delay,  $\tau$ , for which the correlation coefficient,  $C_C = \max\{c(\tau)\}$ , occurs measures the delay between the two seismograms (Fig. 1c). The Fourier transform of  $c(\tau)$  is equivalent to the cross spectrum

$$C(f) = n \langle U_1(f)U_2^*(f) \rangle. \quad (2)$$

Its absolute amplitude,  $|C(f)|$ , after smoothing across adjacent frequencies, is called the coherence, and the delay,  $\tau$ , can also be estimated from the slope of the phase (Fig. 1b, d), for example, in a regression weighted by the coherence.

Directly measuring arrival time differences by cross correlation requires the choice of windows around previously specified approximate arrival times for a pair of events. These must be estimated visually (catalog pick), by an automatic detector, or, alternatively, as a theoretical arrival time if the location is roughly known. In each case the window may be grossly misaligned and contain only a fraction of the true similar waveform needed for a useful comparison. We have encountered this problem even for the nearly identical waveforms of repeating events, where a particularly noisy day at a station may obscure the first arrival for one of the events. For this reason, we need to be able to recover initial window offsets as large as approximately 1 sec (100

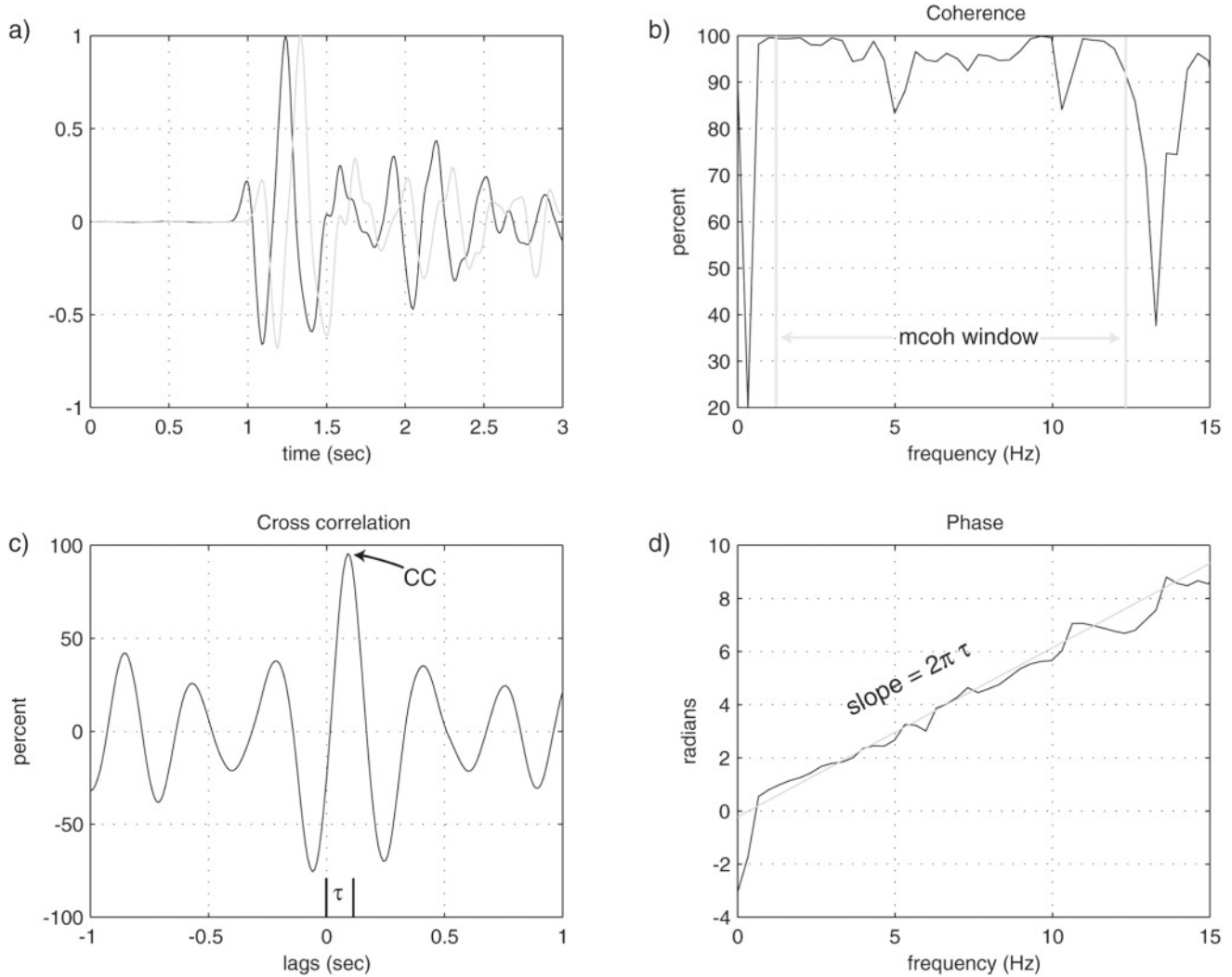


Figure 1. (a) Principle of delay estimation for two traces using (c) time-domain cross correlation and (b, d) the cross spectral approach. Similarity measures are cross correlation coefficient ( $C_C$ ) and mean coherence. The computed lag is  $\tau$ .  $CC = C_C$  and  $mcoh = m_{coh}$ .

samples for the NCSN) to maximize the total number of observations.

Time-domain cross correlation and the cross spectral approaches differ substantially in their performance in recovering large delays. This is shown in Figure 2, which illustrates an attempt to recover lags of up to 1 sec for identical, noise-free seismograms using 256 sample time windows for a variety of offsets. The similarity measure used in the time domain is the correlation coefficient ( $C_C$ ) and in the frequency domain the mean coherence ( $m_{coh}$ ), the latter taken over a limited frequency band from 2.5 to 12 Hz. The lag,  $\tau$ , should be the diagonal line  $y = x$ , and the similarity measure should equal 100% for each run. The cross spectrum technique shows a marked decay in  $m_{coh}$  with increasing delay due to the necessary averaging in the complex domain. Such an averaging over a series of complex numbers with rapidly varying phase biases the result toward zero

(Jenkins and Watts, 1968). Note that a more sophisticated implementation for estimating the delay from the cross spectrum using only the first few prolate spheroidal eigentapers reduces this tendency (Aster and Rowe, 2000). This and problematic phase unwrapping at large lags lead to a recovery of the true delay for lags only up to eight samples for the simple cross spectrum technique. Clearly, time-domain cross correlation is more robust and reliable for recovering large lags.  $C_C$ , however, is also seen to decrease with increasing lag due to the fact that a smaller fraction of the waveforms are identical in the offset windows. Although an associated large lag may be correct, the observation may be discarded due to a low  $C_C$  value. Therefore we employ a two-step procedure: first align to the nearest sample, then compute to subsample precision and a more accurate  $C_C$  value. A similar methodology was used by Aster and Rowe (2000) and Rowe *et al.* (2002), except that they computed

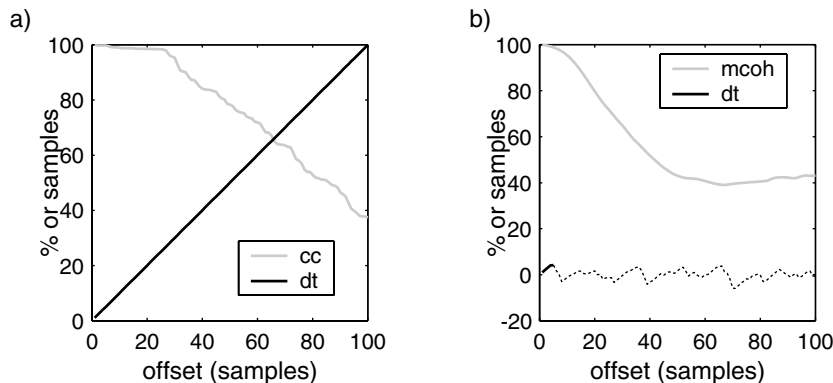


Figure 2. Comparison of the performance for (a) time-domain cross correlation and (b) cross spectral approaches for measuring time lags as a function of offset. The window length is 256 samples.  $CC = C_C$  and  $m_{coh} = m_{coh}$ .

to subsample precision in the frequency domain, whereas we fit a parabola in the time domain in the vicinity of the peak of the cross correlation function. For the first step we both compute the integer sample cross correlation function by multiplying in the frequency domain and then taking the inverse Fourier transform.

We chose a time-domain cross correlation approach since root mean square (rms) misfits are slightly smaller than for the cross spectral approach and computations can be significantly faster. We also employ window lengths of 256 samples to retrieve as many large offsets as possible and to obtain the most observations for the relocation. Longer windows have the added benefit that they have a greater chance of capturing coherent energy if the waveforms are initially misaligned or if the signal-to-noise ratio is low.

### Residual Analysis and Observation Criteria

Relocating earthquakes using relative arrival times requires measurements that connect earthquakes with other nearby events. This is feasible for closely spaced earthquakes with similar waveforms (discussed later). It becomes more problematic as waveforms become less similar, for example, due to larger source separations. Differences in earthquake mechanism, heterogeneous Earth structure, and effects from source finiteness also contribute to waveform dissimilarity.

To obtain precise locations for the entire earthquake catalog, we seek to maximize the number of reliable correlation observations. This can be achieved in two ways. First, the correlation function itself can be made more robust to provide better measurements, as discussed in the previous section. Second, as described in this section, threshold criteria may be relaxed so as to allow more measurements to be retained. We have used two measures of similarity, the normalized cross correlation coefficient ( $C_C$ ) and the mean coherence ( $m_{coh}$ ), and we compare their performance later. Not every correlation measurement yields useful information for the relocation. Less reliable data and outliers with unacceptably large errors need to be excluded. After we have iterated the linearized inversion process and converged on a solution, it is also possible to use the size of the residuals to

identify inconsistent and therefore probably less reliable measurements.

Figure 3 displays a connectivity plot as an example of interconnectedness for a set of earthquakes using only correlation data. This data set includes 1494 relocated events on the Calaveras Fault where the observation criteria are  $m_{coh} > 90\%$  and  $C_C > 70\%$ . The connections between earthquakes in this figure demonstrate that many of these events could be relocated with correlation data alone. These events represent about 95% of the seismicity in the catalog for this section of fault. It is not clear, however, how reliable the measurements underlying these connections are compared with the catalog data over these length scales and ranges of magnitudes.

For this purpose, we use the double-difference relocation program (Waldhauser and Ellsworth, 2000) and analyze the residuals of the relative arrival times after relocation. The residuals convey information on how errors vary with cross correlation coefficient, mean coherence, interevent distance, and magnitude. Equating the postfit residuals with the size of the measurement error assumes that the model on which they are based (velocity model and travel-time calculation) is correct. This is never strictly true, and hence the residuals will not entirely characterize the true error. However, we can estimate the measurement error from repeated measurements, obtained from sequences of repeating earthquakes. In the following, we focus on the total error due to both measurement error and unmodeled velocity structure, which we assume is identical to the postfit residuals for this overdetermined problem.

Figure 4 illustrates the distribution of 926,571  $S$ -wave residuals as a function of  $C_C$ . These arise from a single location run for the 1494 events on the Calaveras Fault using  $C_C > 70\%$ . We observe more scatter (represented in terms of the standard deviation of the residuals) for lower values of the cross correlation coefficient.

Analyzing residuals is advantageous because the residuals map a variety of parameters with different units to a common system in which they may be compared directly. For instance, the nominal cutoff of 90% often employed for  $C_C$  and  $m_{coh}$  has no direct meaning in terms of location.

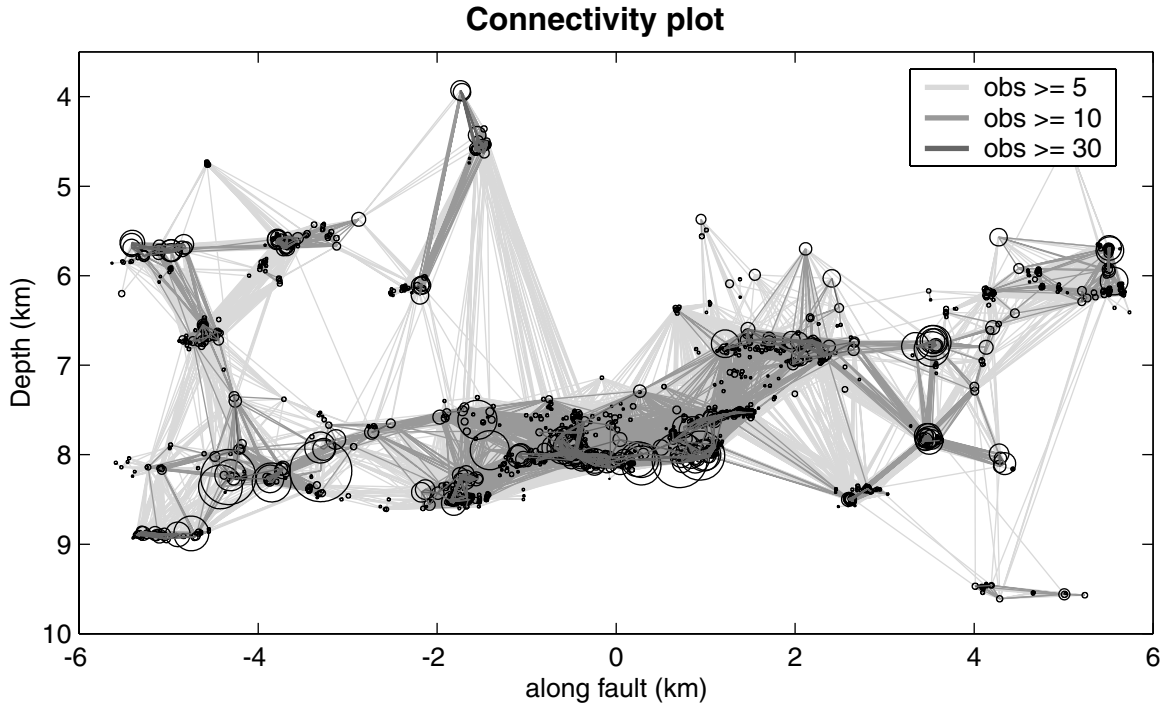


Figure 3. Constraint from correlation data with  $m_{coh} > 90\%$  and  $C_C > 70\%$  for 1494 events on the Calaveras Fault. If an event pair has more than 30 observations, it means more than 30 stations match the given criteria.

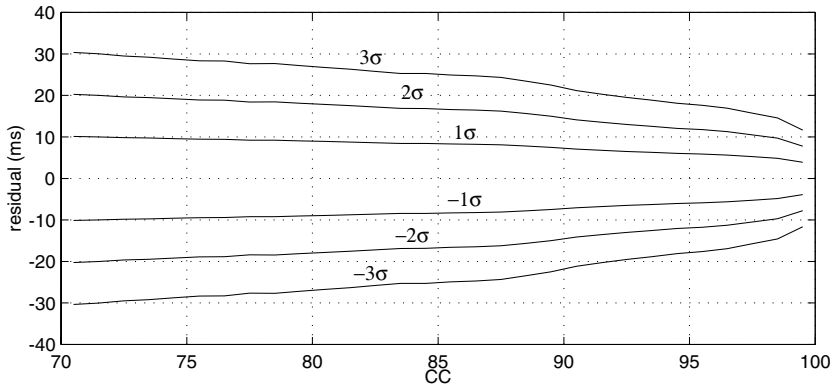


Figure 4. Distribution of 926,571 S-wave differential travel-time residuals from the 1494 Calaveras Fault earthquake relocations as a function of correlation coefficient. Contours shown for integer standard deviations.  $CC = C_C$ .

Residuals in units of milliseconds are more directly tied to location uncertainty.

Figure 5 shows the standard deviations of the residuals and the number of observations as a function of both  $C_C$  and  $m_{coh}$  for the 1494 events on the Calaveras Fault. Note that in Figure 5b, the maximum ridge of the surface is shifted to the right of the diagonal line  $C_C = m_{coh}$ . This indicates that  $m_{coh}$  is a less stringent measure of similarity; the threshold value of 90% (shown by a gray line) would accept 275,000 measurements as observations for  $m_{coh}$ , whereas  $C_C$  would allow only 160,000.

We expect our most reliable measurements to have both high  $C_C$  and  $m_{coh}$ , which is substantiated by the location residuals (Fig. 5a). Many observations, however, with 90%  $m_{coh}$  have only 70%  $C_C$  and therefore are somewhat suspect.

This is confirmed by their higher postfit residuals. We choose a double selection criterion,  $C_C + m_{coh} > 170\%$ , since it captures most of the measurements that have standard errors at or below the sample rate (10 msec). This criterion was employed in the study of Schaff *et al.* (2002) based on these initial tests on a subset of the Calaveras Fault seismicity. If only a single parameter is desired,  $C_C$  provides a more robust means of identifying good-quality observations than does  $m_{coh}$ . For example, at the 90% level, the standard deviations of the residuals are  $\sim 10$  msec for  $C_C$ , whereas for  $m_{coh}$ , they range from 5 to 30 msec. In other words, the similarity measure  $C_C$  corresponds more closely to the location prediction error.

Applying both criteria, we obtain many more observations (600,000) than by applying either one alone at the 90%

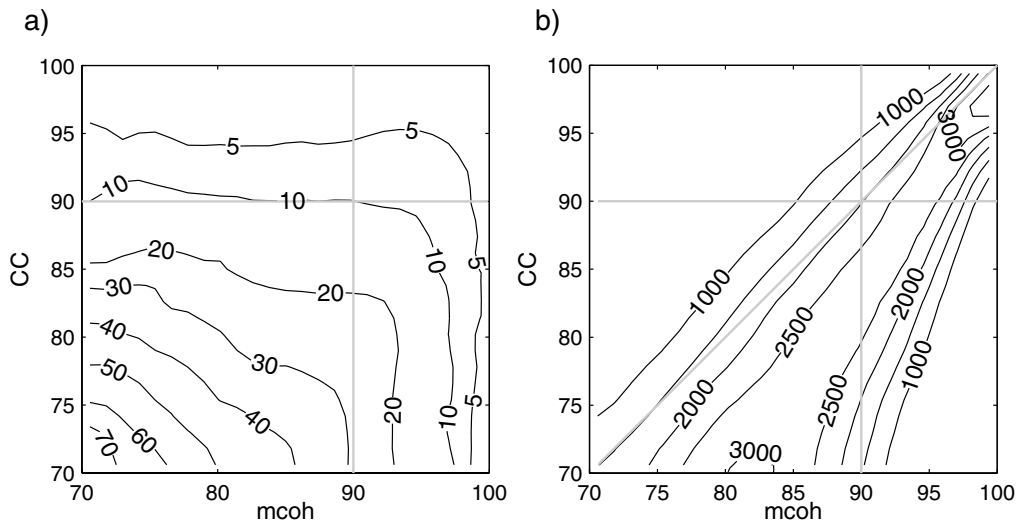


Figure 5. (a) Standard deviation of residuals (msec) for the 1494 events as a function of similarity measures. (b) Number of data measurements per unit bins of  $C_C$  and  $m_{coh}$ .  $CC = C_C$  and  $mcoh = m_{coh}$ .

level. Previous studies have used more strict thresholds, from 80% to 90% (e.g., Poupinet *et al.*, 1984; Got *et al.*, 1994; Dodge *et al.*, 1995; Rubin *et al.*, 1999; Waldhauser *et al.*, 1999). We find that for our application, those selection criteria are overly conservative, and we instead pose the practical question of when the correlation measurements are better than the corresponding catalog phase data. For example, catalog picks will never be more precise than the sample rate (10 msec in this example). We find that much lower thresholds may be applied for the purposes of earthquake location, provided weighting of the observations is carefully done.

### Estimating Measurement Error

A reasonable way to estimate correlation measurement error for a cluster of similar events follows from the internal consistency of the measurements. The measurements provide a constraint on an overdetermined set of equations ( $t_{13} = t_{12} + t_{23}$ , where  $t_{ij} = t_j - t_i$ ). However, this condition is not sufficient to estimate true measurement error of the correlation data, as we illustrate in a simple example in Figure 6. Consider two multiplets recorded at one station, as seen in Figure 6a. Internally, their waveforms are very similar, but between the multiplets they are dissimilar, as is reflected by the correlation coefficient matrix in Figure 6b with  $C_C \sim 50\%$  on a 50-sample-long window. The corresponding delays in Figure 6c for these low-coherence areas are consistent, but they are in error. This is because of their opposite polarities; the troughs are aligned by cross correlation, causing the first breaks to be offset by about six samples, in accordance with the delay matrix. Based on the internal consistency of these measurements, the error would be incorrectly estimated as quite low. This problem occurs when

inverting at only one station and arises from the difficulty in defining appropriate similarity clusters. By performing the inversion at several stations (i.e., solving for the location), the biases between clusters will be inconsistent, which will be reflected in the residuals. This approach captures more of the true measurement error than measuring the internal consistency, but it relies on the assumptions that the problem is overdetermined and that the locations are well constrained.

By estimating the measurement error from location residuals, we are also able to compare two very different data types and discriminate under which conditions one is better. The residuals are computed based on the best final locations. It is unlikely that they depend strongly on the details of the location algorithm used. The double-difference technique is our method of choice because of its advantages in reducing velocity model error and inclusion of correlation data to determine locations.

Another advantage of the double-difference algorithm is that there is no need to define clusters of events. We want to estimate errors as a function of event separation distance and magnitude, which would be more difficult if events had to be clustered. Finally, there is no way to estimate catalog pick error based on internal consistency at a single station. The differential arrival times are perfectly consistent, implying zero error, which is false.

### Interevent Distance

Correlation measurements, unlike arrival time pick measurements used in catalog locations, are known to degrade with interevent separation distance. This is because waveform similarity breaks down due to increasingly different earth structure sensed by waves from more widely separated events. A Fresnel zone argument suggests that

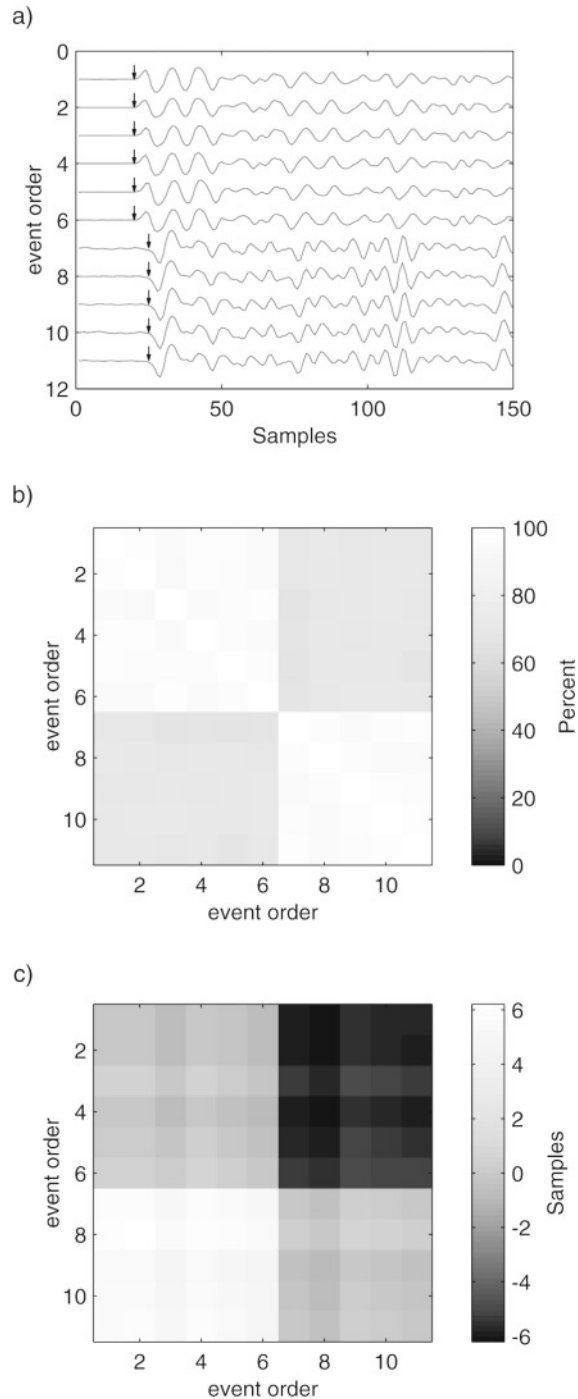


Figure 6. (a) Two multiplet clusters aligned by correlation, true arrival times reflected by arrows. (b) Corresponding  $C_C$  matrix and (c) delay matrix. Similarity between clusters is low, but delays are all consistent. The peaks and troughs from the opposite polarities have become aligned, causing a biased half-cycle offset between the two clusters.

waveform similarity should hold at separation distances of up to a quarter of a wavelength (Geller and Mueller, 1980). More uncertain is the reliability of correlation measurements for kilometer-scale separation distances, as seen in Figure 3.

To examine this question, we consider a subhorizontal streak of earthquakes from  $-1$  to  $1$  km along strike at  $8$  km depth for the Calaveras Fault seismicity illustrated in Figure 3. Figure 7 shows the location results for the streak using catalog phase data only (Fig. 7a) and correlation measurements only (Fig. 7b). For both calculations, the same 299,642 observations are employed (identical event pairs and stations). In the latter case these are measured directly from correlation data and in the former case calculated from arrival time differences of visually estimated picks (catalog data). In this way, relative measurement error between the two data types can be compared. Every other aspect of the inversion is the same. Although still unknown, the velocity model error is fixed and common to both. To obtain the most matching observations for the catalog data, we use all correlation measurements ( $C_C > 0$ ). The window length is 256 samples (2.56 sec). Qualitatively, the correlation data spanning length scales between  $0$  and  $2$  km and for this range of magnitudes ( $0-3.5$ ) yields a much better definition of the streak.

Figure 8 examines more quantitatively the effect of event separation on the measurement error by plotting the median and standard deviation of the residuals as a function of separation distance. Each statistic is computed for bins of 12,000 observations. Extreme outliers are removed if they differ from the mean by more than 10 times the median absolute residual. At near-zero separation distances, the residuals reflect only measurement error for the two data types in the ideal case, which amounts to approximately 3 msec for the correlation measurements and approximately 11 msec, or just over one sample, for the catalog measurements (Fig. 8a). Thus the correlation data are fit much better than the catalog data. In fact, the latter cannot be more precise than 10 msec, the sampling interval. The increase in residuals with distance for the catalog data is due to model error, that is, uncertainty in the travel-time predictions from the earth model. For the correlation data, it is a combination of both increased measurement and increased model error with separation distance. Note, however, that the increase in residuals for the cross correlation data with increasing event separation is quite slow, at a rate comparable to the increase for the catalog data. This indicates that the increase in residuals for the cross correlation measurements is due primarily to unmodeled velocity structure, rather than measurement error. The average residuals from the correlation measurements for this streak are at least a factor of 2 smaller than the catalog data up to separation distances of 2 km.

We also have some independent information on the quality of each observation for both the catalog and correlation data. For the catalog observations, this information consists of a subjective quality measure ranging from good

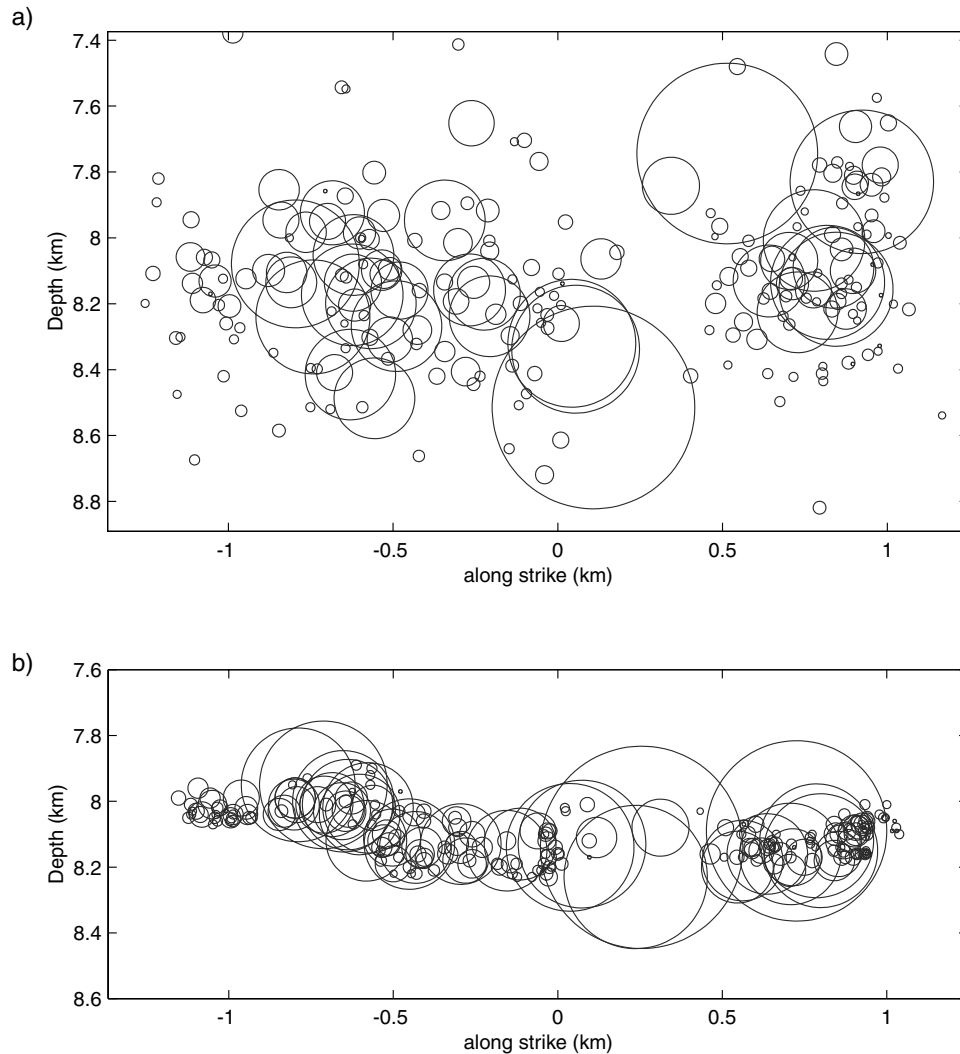


Figure 7. Comparison of measurement error between (a) original catalog only and (b) correlation only data consisting of 299,642 identical observations (model error fixed) for the relocation of a 243-event streak.

(0) to poor (3). For computed differential times, the average weight can thus assume half-integer values. For the correlation measurements, we use the cross correlation coefficient as a measure of measurement quality. Figure 8c shows that this prior information largely predicts the level of error as expressed by the median residuals. At zero separation distance, the residuals reflect measurement error alone and indicate why the best correlation data ( $90\% < C_C < 100\%$ ,  $\sim 2$  msec) result in nearly an order-of-magnitude improvement over average catalog data (weight 1.5,  $\sim 18$  msec). It is surprising, however, that correlation measurements with correlation coefficients as low as 40% still appear better than the best catalog data for this streak, as reflected in the residuals. This typically occurs when a coherent signal exists under poor signal-to-noise conditions, as has been seen for the case of repeating events.

### Magnitude

The same approach may be used to compare the errors in correlation and catalog data as a function of magnitude. In Figure 9a, correlation measurement quality decreases with increasing magnitude, as indicated by increasing residuals. This most likely due to the effects on seismograms of clipping and possibly also to source finiteness. For the catalog data, however, measurement quality increases slightly for the larger events because of a better signal-to-noise ratio. Still, the correlation data achieve subsample precision and thus outperform the catalog data for this range of magnitudes (0–3.5). Moreover, the correlations are able to recover many more differential time measurements for smaller magnitudes (by initially aligning on theoretical arrivals), for which only a few catalog picks are available due to low signal-to-noise

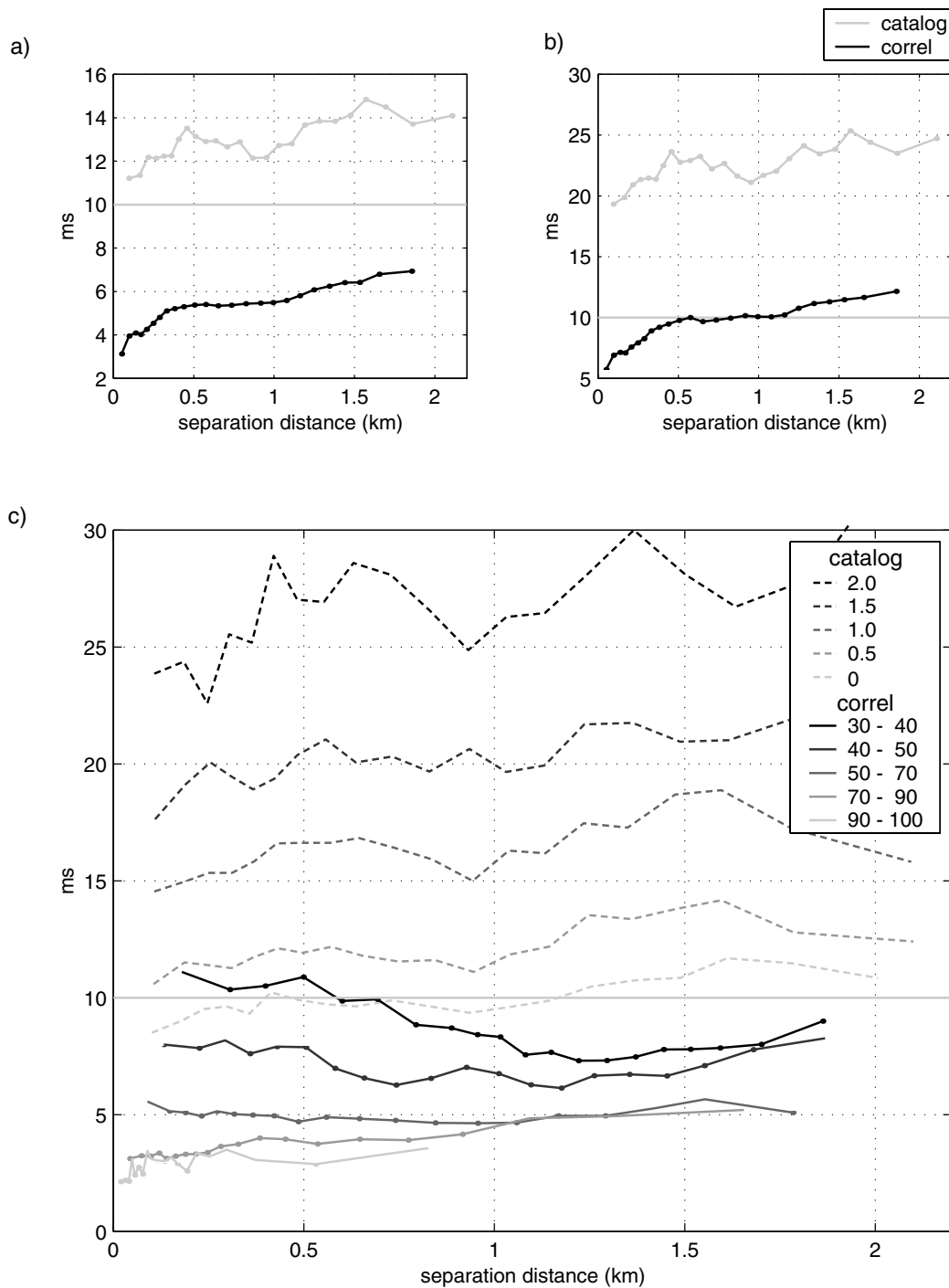


Figure 8. Statistics for 299,642 differential travel-time residuals (243 events). (a) Catalog and correlation median absolute residuals shown as a function of interevent separation distance for 25 bins each containing 12,000 observations. (b) Same, but for standard deviation. (c) Median residuals separated according to quality. Lines are shown for catalog data with weights 0, 0.5, 1, 1.5, and 2 and for correlation data with  $C_C$  values between 30 and 40, between 40 and 50, and so on.

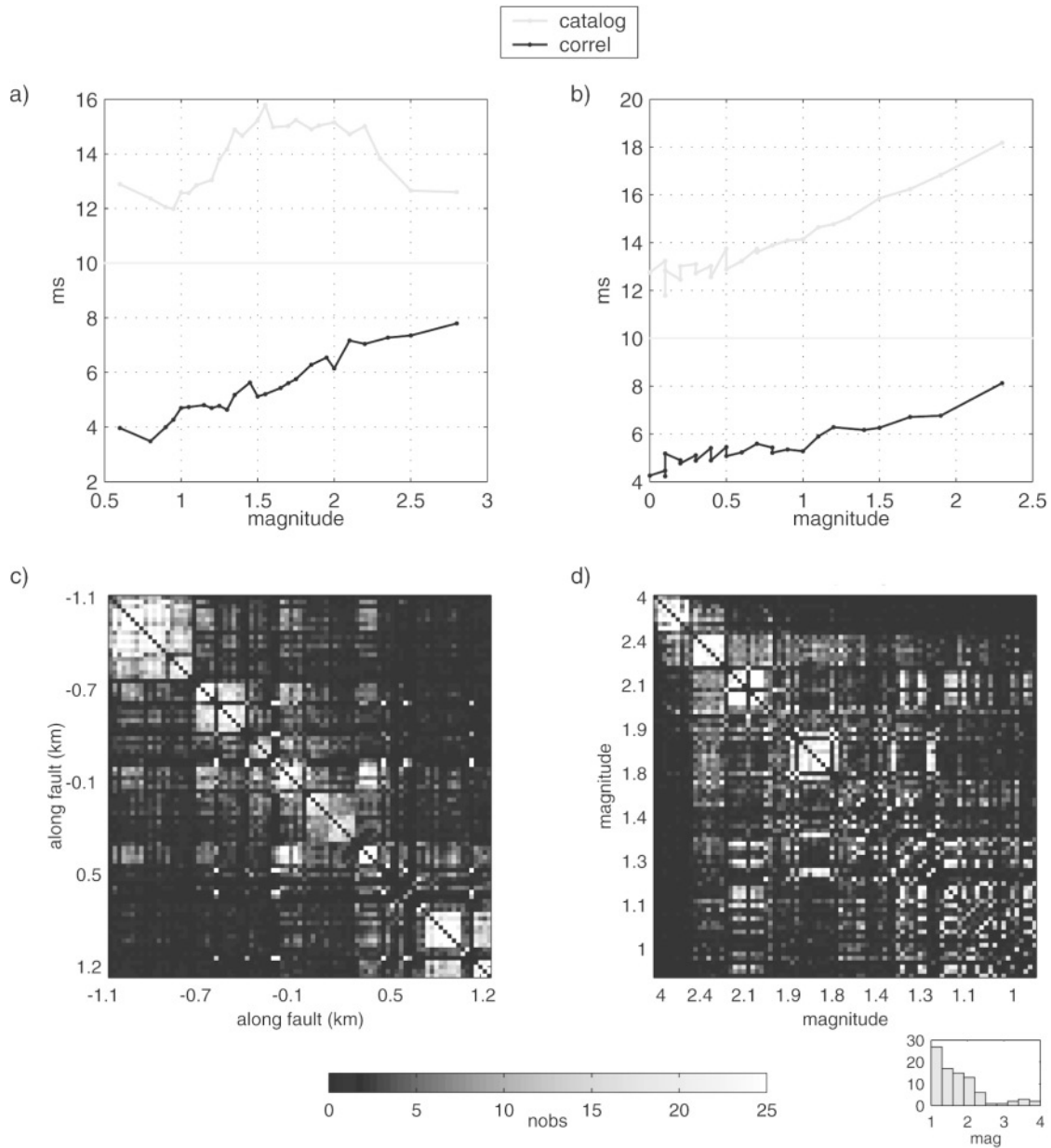


Figure 9. (a) Median absolute residuals for 25 bins each containing 12,000 observations as a function of mean magnitude for the 299,642 event pairs (243 events). (b) Median residuals as a function of magnitude difference. (c) Observation matrix for 88 events at Bear Valley ordered by distance along strike. (d) Same matrix ordered by magnitude. The observation matrix is related to the correlation matrix but refers to the number of stations that meet the observation criteria for each event pair.

ratios. Cross correlation measurements remain more precise than the catalog for the larger events, although the cross correlation coefficient, which we use to assess measurement quality, decreases with increasing magnitude. This is primarily a result of waveform clipping. The measurements remain precise in part because clipping at these levels of ground motion is due to the limitations of the telemetry system, rather than true clipping of the sensor (Ellis and Lindh, 1976). Thus, some aspects of the true ground motion, such

as the zero crossings, should be preserved with fidelity and hence may continue to give reliable measurements. Rubin *et al.* (1999) noted that most of the unrelocated events in their study were either small- or large-magnitude events when they used  $C_C$  values greater than 85%–95%. Our use of lower waveform similarity enables us to extend cross correlation to larger events. If thresholds are set too high, many useful observations may be discarded.

Figure 9b displays median residuals as a function of the

difference in magnitude for event pairs. We expect larger events not to correlate as well with small events, which the trend reflects. The catalog residuals also increase with magnitude difference, suggesting that the different signal-to-noise ratios between the two events introduces error comparable, in a relative sense, to that introduced into the correlation measurements by waveform differences (see also Fig. 9a). We find that large events within the streak correlate well with other large events despite minor differences in location, clipping, and finite-source effects. Evidence to this effect is also shown in Figure 9c,d, which shows the number of observations (observation matrices) for a 2-km-long streak of earthquakes at Bear Valley. The 88 events are ordered on both axes in Figure 9c as function of distance along strike and in Figure 9d as a function of magnitude. Over this distance scale, events larger than magnitude 1.8 seem to correlate best with other events of like magnitude, rather than with smaller events in closer proximity. In fact, the large events often do not even have overlapping source areas. For smaller events, interevent distance seems to be the controlling factor (Fig. 9c).

We have presented two lines of evidence, the sharpness of the relocation results and the lower residuals for correlation measurements, to demonstrate that correlation data can be significantly better than catalog data for relocations using travel-time differentials, over greater distance and magnitude ranges than previously assumed. The waveforms themselves also support this conclusion. Figure 10a shows the waveforms for the streak shown in Figure 7 at a nearby station (CCO) aligned on the catalog  $P$  picks. Figure 10b shows the same seismograms aligned by cross correlation on the high-energy  $S$ -wave arrival between 450 and 520 samples. The events are ordered along strike from northwest to southeast. This is essentially a receiver gather of the kind commonly used in controlled source seismology. The move-out for the  $P$  wave coming in at around 100 samples is consistent with the 2-km length of the streak. Several features are interesting to note. Visually, the similarity extends over the 2-km range for the correlation-aligned data, and it is a clear improvement over the catalog data. The largest events appear as black and white horizontal bands because they are clipped, but since the phase is preserved when the amplitude is clipped due to the telemetry, the zero crossings allow the cross correlation to successfully align these events. The magnitude 3.5 event corresponds to trace 119. The seismograms from 30 to 42 have aligned well and show magnitude 3's and 1's interspersed. Here is evidence that even events of substantially different magnitude correlate sufficiently well to be aligned properly. These same events in the upper plot (Fig. 10a) appear to have interchanging polarities from 450 to 520 samples because they are shifted (mispicked) by up to half a cycle, roughly a 20-sample error. Because the correlation data are aligned to the nearest sample or better, this translates to greater than an order-of-magnitude reduction in the measurement error.

## Window Length

Cross correlation measurements are fundamentally different from first-break measurements of arrival time in that they align a packet of energy of some duration. Thus, an important parameter in measuring arrival times using cross correlation is the window length over which the seismograms are aligned. Figure 11 compares 869,345 identical observations for different window lengths for the streak in Figure 7. As might be expected, there is a strong tendency for decreasing correlation coefficients with increasing window length (Fig. 11a). As a consequence, the number of observations decreases with increasing window length for a set observation threshold. However, the rms residuals are smaller for larger time windows (Fig. 11b). In terms of residuals and the sharpness of the relocations, the longest window length of 256 samples gives the best results for the streak. It implies that the slowness of the arrivals at the source is similar to that of the direct arrival over the entire window considered. This was previously observed for a similar data set and suggests that much of the early coda is generated near the site rather than over a larger scattering volume (Dodge and Beroza, 1997). If the observation criteria thresholds are too high, many of these better measurements may be discarded. Longer windows seem to work best overall for the regions we have examined. This observation may be somewhat biased, however, compared to other seismic source regions, because of the apparently anomalous abundance of repeating events and streaks for the areas discussed herein. A more detailed residual analysis, as presented earlier, could help to determine whether longer windows are valuable for greater separation distances as well.

## Empirical Weighting Functions

Residual analysis has provided useful guidance in choosing appropriate thresholds for the similarity measures,  $C_C$  and  $m_{\text{coh}}$ , and has helped to justify the application of correlation data to larger separation distances and higher magnitudes. The residuals may also be used to develop empirical weighting functions for the location algorithm to refine the results for a region of interest. Taking the inverse of the standard deviations of the residuals is a natural choice for weighted least squares.

There are two types of weightings currently used in the double-difference location algorithm: *a priori* and dynamic (Waldhauser and Ellsworth, 2000). The weighting functions are theoretical and cutoff parameters are typically determined on a trial-and-error basis. Biweight functions are used to decrease rapidly the interevent weighting for greater separation distances, because it is assumed that travel-time error increases due to an inaccurate velocity model. Empirical weighting functions may be applied after the best locations have been obtained using theoretical functions, to optimize the locations for a particular region. It is important for a stable inversion to use the theoretical weighting functions

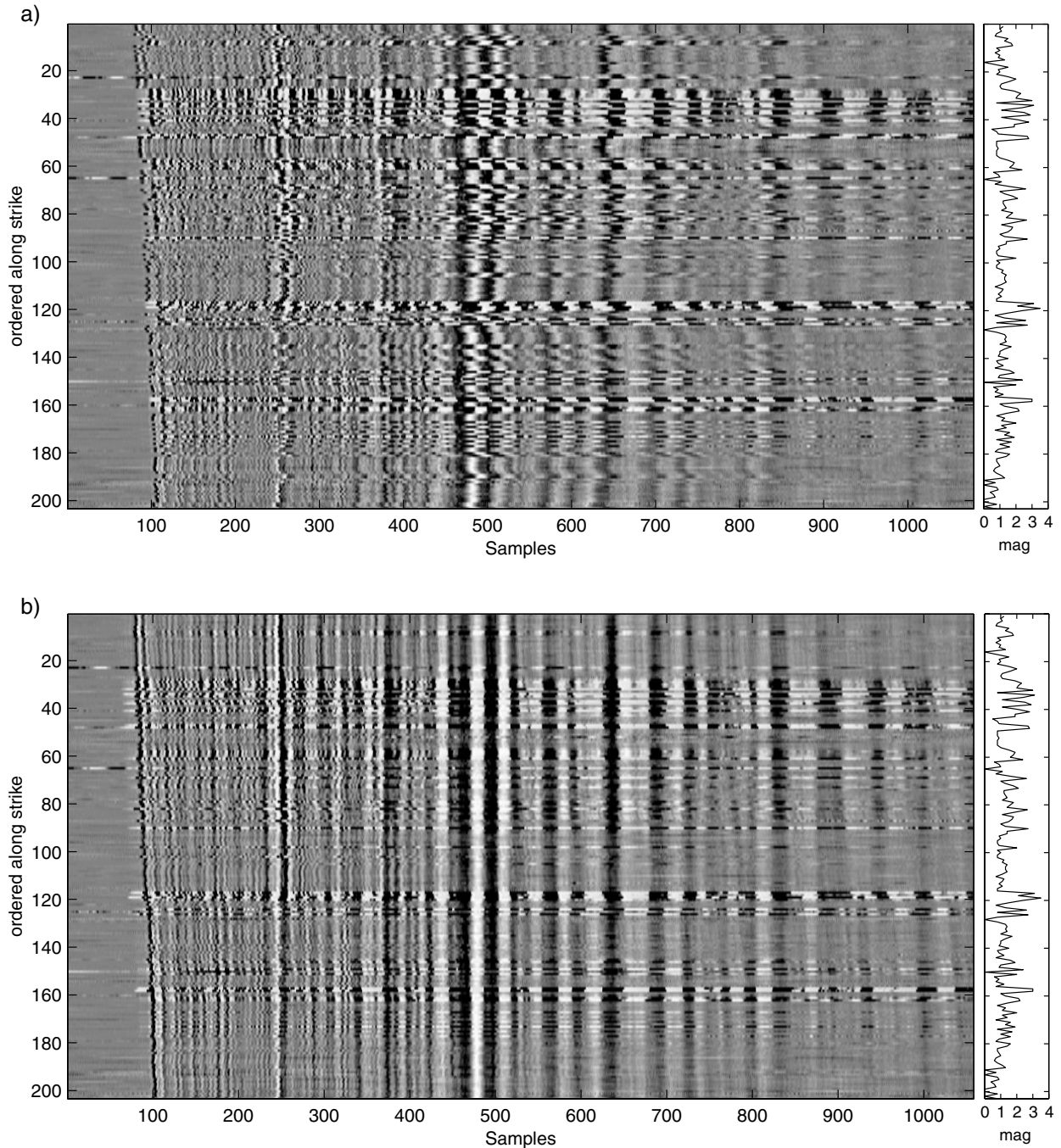


Figure 10. Unfiltered seismograms at station CCO for the 243 event 2-km streak ordered by distance along strike (a) aligned according to catalog *P*-wave picks and (b) aligned by cross correlation on the *S* waves (see text). Images show a 3D surface where peaks are light and troughs are dark. Similarity across the entire streak supports the notion that useful correlation based arrival time measurements can be obtained over these distance scales.

first, so that the locations do not heavily depend on a circular weighting scheme.

After the best locations have been derived, the postfit residuals provide information about the distribution of errors in the data and modeling. The approach currently taken in hypoDD to achieve the best initial locations using mixed

catalog and correlation data seeks to exploit the strengths of the two different data types (Waldhauser, 2001) and is summarized as follows. The first several iterations place all the weight on catalog phase picks to remove the influence of velocity model error over long distance ranges. As the locations improve and interevent distances become more ac-

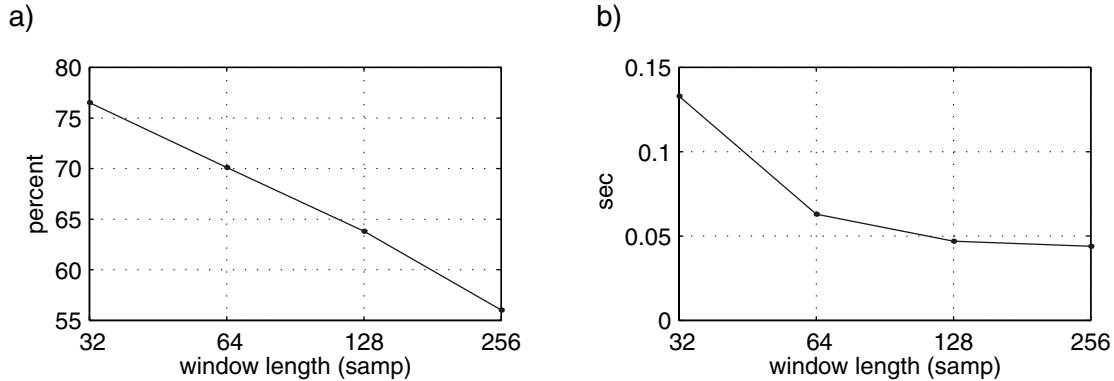


Figure 11. The 869,345 identical observations for different window lengths for the 243-event streak. (a) Longest windows (256 samples) have lower correlation coefficients but (b) produce sharper locations and better rms misfit for the streak.

curate, the correlation data are given increasingly greater weight because of their superior measurement precision for shorter distance ranges.

Because we observe that *a priori* parameters such as  $C_C$  can be mapped into the postfit residuals, we can better characterize the errors for the inversion by converting the similarity measurement into an empirical weighting factor to be used for subsequent runs. The standard deviations obtained as a function of  $C_C$  (as in Fig. 4, for example) may be used for this purpose (Fig. 12a). Figure 12c compares an empirical weighting function calculated from Figure 12a as  $\sigma^{-2}$  (normalized from 0 to 1) with a linear weighting function and the commonly used Hannan–Thomson weighting function. Since this function is likely to vary from region to region, an appropriate scheme may be derived by first iterating on the locations, performing the residual analysis, and then relocating using a more appropriate weighting for the region under study.

The second type of weighting is applied dynamically during the inversion. Perhaps most important is the inverse weighting with interevent distance. For catalog phase data, this is essential to accommodate model error over larger separations. It is also important, however, for correlation data since the delay measurement itself also becomes less reliable with increasing distance. Figure 12b,d illustrates the conversion of residuals to distance weighting functions for both catalog and correlation data for the 1494 events in Figure 3. As noted earlier, residual analysis has the useful property that it can compare very different aspects of a problem in a common frame of reference. Mapping everything to units of time allows for direct, quantitative comparisons. For example, Figure 12b suggests that catalog weight zero and correlation data should probably be given equal weight at about 500-m separation distance. At zero separation, the correlation measurements should have 5 times the weight of the best catalog measurements. Beyond 2 km, the correlation data are sparse and have substantially more error than weight 1 data from the catalog, so perhaps those observations should be removed. Note also for this data set of 1494 events that

correlation data are not better than catalog phase data at distances up to 2 km, in contrast with the example of Figure 8c; this confirms that optimal weighting functions for different sets of seismicity will vary.

This also raises the issue of thresholds versus weighting functions. A benefit of weighting functions is that no hard cutoff needs to be determined, and so concerns of not having enough connectivity or observations are reduced. Weighting functions assign a unique value to each data point from a continuous spectrum according to its error. The value of thresholds, on the other hand, is that they reduce the size of problems that might otherwise be too large, so that only the best data are retained. Thresholds also eliminate severe outliers, which could unacceptably degrade a solution.

### Spatial and Temporal Variability of Waveform Similarity

The success of waveform correlation measurements in providing improved data for earthquake location and related applications depends entirely upon the similarity of the waveforms (low measurement error). We have demonstrated a quantitative means of assessing this improvement and have extended correlation measurements to less similar events. It is useful, however, to understand the factors that affect the similarity of waveforms. As discussed earlier, waveform similarity depends upon interevent separation distance. More completely, though, it also depends upon differences in the 3D seismic phase velocities, when ray paths diverge. It may also change temporally if the properties of the medium change over time. Moreover, waveform similarity is also reduced when source properties differ. Colocated sources may give rise to dissimilar waveforms if the source mechanisms are different, and, for large events, waveforms arising from dissimilar source-time functions will naturally also differ.

Figure 13 displays the  $C_C$  matrices for seven nearby clusters of repeating events and how these matrices vary by station, depending on station location. The order of the  $C_C$

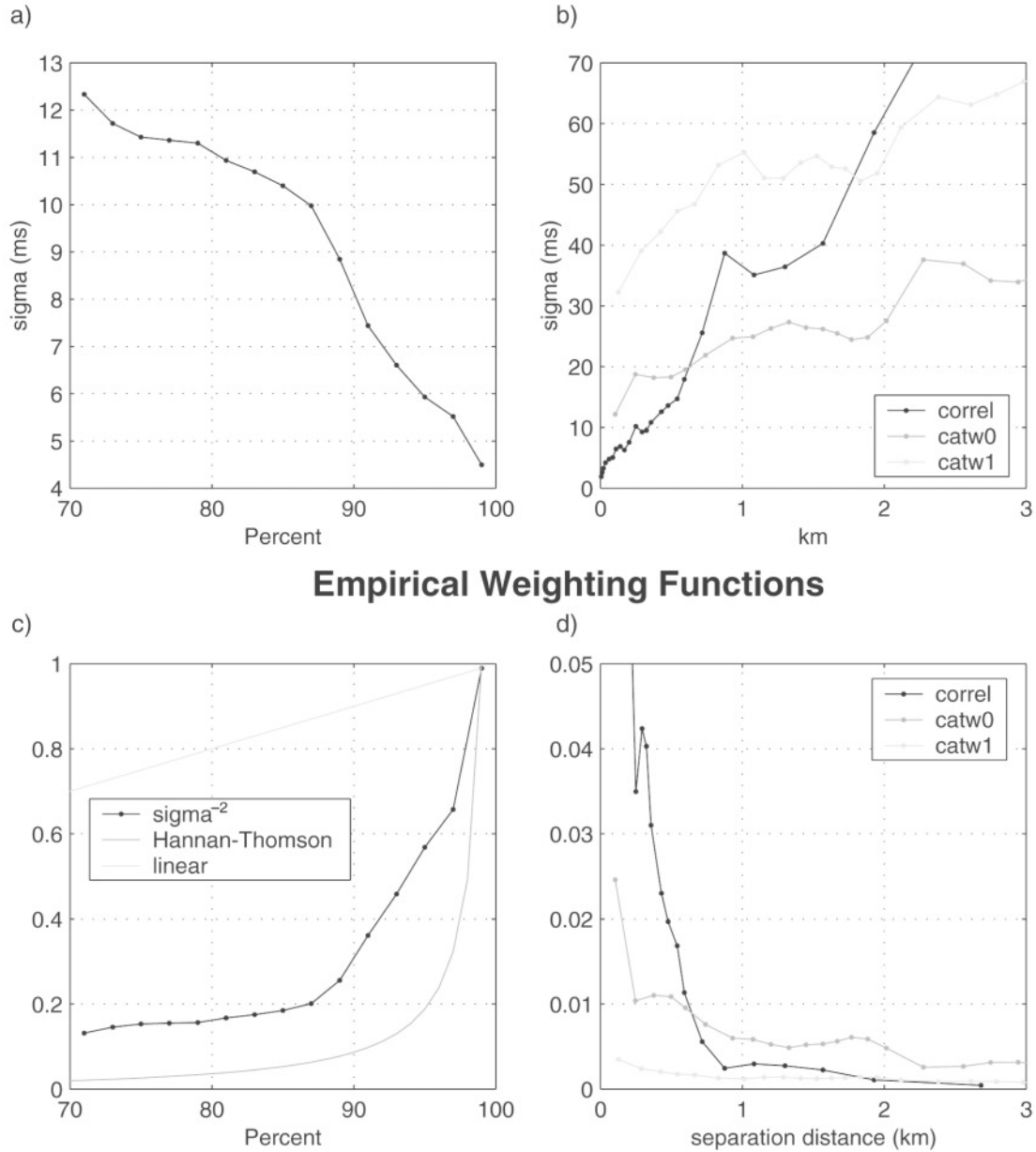


Figure 12. (a) Standard deviation of 926,571 residuals (1494 events) as a function of  $C_C$  (same as Figure 4). (b) Standard deviation of residuals as function of interevent distance for different data types. (c) Example weighting functions based on data quality. (d) Empirical weighting functions taken as the inverse variance normalized from 0 to 1.

matrix is arranged so that the clusters appear as highly similar white blocks on the diagonal. At some stations, such as HCB, there is little waveform similarity between clusters. However, at other stations, such as HKR, most event-pair combinations have high  $C_C$  values. In other words, waveform similarity does not depend only on a simple 1D separation distance relation, or quarter-wavelength rule (Geller and Mueller, 1980). It is also affected by the degree and distribution of velocity heterogeneity between the paths traveled by seismic waves for an event pair. In a homogenous half-space, we would expect waveform similarity to have no dependence on event separation, whereas waveforms tra-

versing a region with a very complex, heterogeneous velocity structure should correlate well only for events with very small separation distances, such that their travel paths are largely similar.

Seismic properties of the medium may also change with time. These may manifest themselves as changes in the waveform over time, including velocity changes, a breakdown in coherence, or attenuation (Baisch and Bokelmann, 2001; Schaff, 2001). We illustrate in Figure 14a the effects of the  $M$  7.1 Loma Prieta earthquake on the waveforms of a repeating event sequence on the San Andreas Fault ordered in time. Note that the pre-Loma Prieta event (the first) is

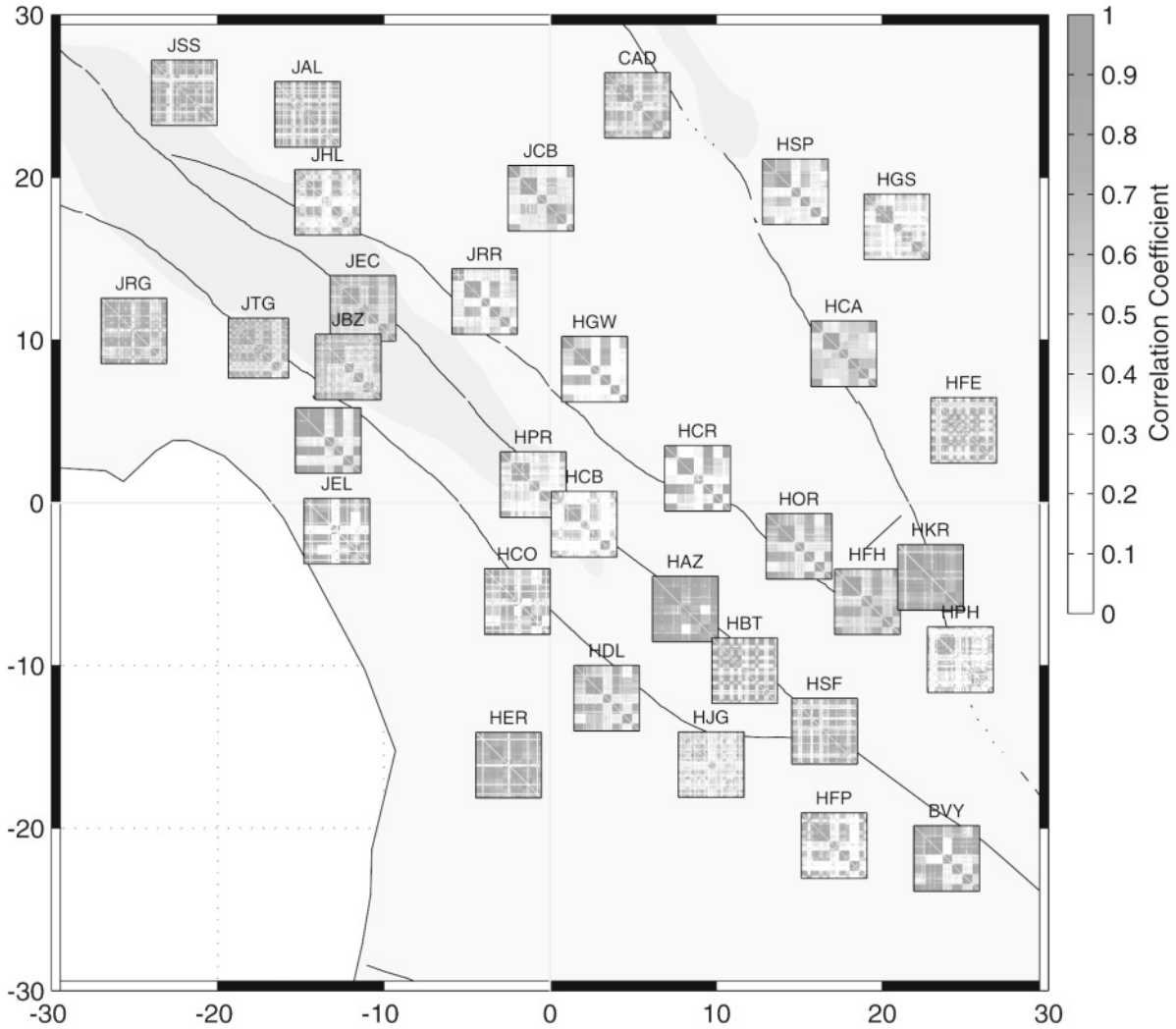


Figure 13. Spatial variability for the  $C_C$  matrix of seven multiplet clusters located at the origin. Multiplets can be discerned as highly similar blocks along the diagonal, at HGW for example. This variation between stations poses a challenge in determining a fixed threshold for the observation criteria applied across all stations. Map axes in kilometers.

most similar to event 14, the last event in the sequence, and is least similar to the event that immediately follows it. This is an effect of time-dependent changes in the medium distorting the waveform due to the effects of the 1989 Loma Prieta earthquake (Schaff, 2001).

### Conclusions

We have presented a general approach to quantify uncertainties in relative arrival time measurements and modeling using postfit residuals. This approach allows us to set data thresholds and weighting functions and to assess quantitatively the benefits and limitations of correlation measurements compared to catalog data. The scaling relations between waveform cross correlation measurements and interevent separation distance, magnitude, and window length

are discussed for a local network (NCSN) and for signal frequencies from 1.5 to 12 Hz; they need not necessarily apply elsewhere. These parameters also depend on the geologic structure of the crust. The appropriate scaling relations may be empirically derived for other regions of interest using the steps outlined here. The process is an iterative one in which postfit residuals can be used from each iteration to fine-tune the results of subsequent runs.

We believe that the combined waveform correlation double-difference scheme can have a similar, 1–2 orders of magnitude, improvement for other regions with dense seismicity, regardless of station geometry, 3D velocity structure, seismicity type, or data quality. These issues are common to all earthquake location problems. The most dramatic improvement in locations arises when both model and measurement error can be substantially reduced. Relative relo-

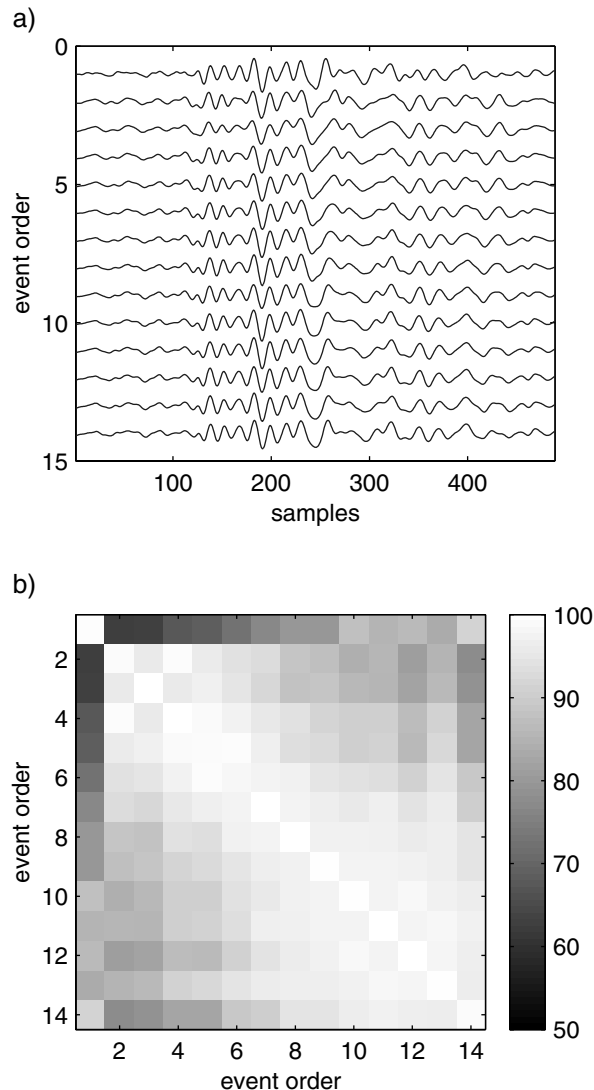


Figure 14. (a) Waveforms for a repeating event cluster that have been altered by the 1989 Loma Prieta earthquake. (b)  $C_C$  matrix reveals the first event (pre-Loma Prieta) is more similar to the last events. Events 2–5 after Loma Prieta, conversely, are more similar to each other (see text).

cation techniques using differential travel times reduce much of the model error that limits absolute travel-time formulations. They are also able to directly invert correlation data, retaining the greatest redundancy using the maximum available observations. When waveforms exhibit sufficient similarity, correlation can achieve subsample precision, which catalog data cannot do. If the input measurements to a location algorithm are improved, the output relocations are sure to be better. It remains to be seen how useful correlation data will be at different spatial scales. Since the waveforms must be similar for correlation measurements to be of use, the method will likely work best where earthquake density is high relative to the length scale of the problem, be it mining seismicity or teleseisms.

## Acknowledgments

We are grateful to Doug Dodge for helpful discussions and development of the codes. We also benefited from conversations with Allan Rubin, Keith Richards-Dinger, and Cliff Thurber. We thank Lisa Block, Charlotte Rowe, Doug Wiens, and an anonymous reviewer for their comments. This work was supported by NSF Grant Numbers EAR-9725238 and EAR-0102803 and USGS Grant Number 00HQGR0062.

## References

- Aster, R. C., and C. A. Rowe (2000). Automatic phase pick refinement and similar event association in large seismic datasets, in *Advances in Seismic Event Location*, C. Thurber and N. Rabinowitz (Editors), Kluwer, Hingham, Massachusetts, 231–263.
- Baisch, S., and G. H. R. Bokelmann (2001). Seismic waveform attributes before and after the Loma Prieta earthquake: scattering change near the earthquake and temporal recovery, *J. Geophys. Res.* **106**, 16,323–16,338.
- Deichmann, N., and M. Garcia-Fernandez (1992). Rupture geometry from high-precision relative hypocentre locations of microearthquake ruptures, *Geophys. J. Int.* **110**, 501–517.
- Dodge, D. A., and G. C. Beroza (1997). Source array analysis of coda waves near the 1989 Loma Prieta, California, mainshock: implications for the mechanism of coseismic velocity changes, *J. Geophys. Res.* **102**, 24,437–24,458.
- Dodge, D. A., G. C. Beroza, and W. L. Ellsworth (1995). Foreshock sequence of the 1992 Landers, California earthquake and its implications for earthquake nucleation, *J. Geophys. Res.* **100**, 9865–9880.
- Dodge, D. A., G. C. Beroza, and W. L. Ellsworth (1996). Detailed observations of California foreshock sequences: implications for the earthquake initiation process, *J. Geophys. Res.* **101**, 22,371–22,392.
- Ellis, J. R., and A. Lindh (1976). Linearity of VCO-discriminator playback system with respect to zero crossing times, *U.S. Geol. Surv. Open File Rept.* 76-873, 8 pp.
- Fréchet, J. (1985). Sismogenèse et doublets sismiques, *Ph.D. Thesis*, Université Scientifique et Médicale de Grenoble, 206 pp.
- Frémont, M.-J., and S. D. Malone (1987). High precision relative locations of earthquakes at Mount St. Helens, Washington, *J. Geophys. Res.* **92**, 10,233–10,236.
- Geller, R. J., and C. S. Mueller (1980). Four similar earthquakes in central California, *Geophys. Res. Lett.* **7**, 821–824.
- Got, J.-L., J. Fréchet, and F. W. Klein (1994). Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res.* **99**, 15,375–15,386.
- Ito, A. (1985). High resolution relative hypocenters of similar earthquakes by cross spectral analysis method, *J. Phys. Earth* **33**, 279–294.
- Jenkins, G. M., and D. G. Watts (1968). *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, 525 pp.
- Neuhauser, D. S., B. Bogaert, and B. Romanowitz (1994). Data access of northern California seismic data from the Northern California Earthquake Data Center (abstract), *EOS* **75**, 429.
- 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.
- Rowe, C. A., R. C. Aster, B. Borchers, and C. J. Young (2002). An automatic, adaptive algorithm for refining phase picks in large seismic data sets, *Bull. Seism. Soc. Am.* **92**, 1660–1674.
- Rubin, A. M., D. Gillard, and J.-L. Got (1999). Streaks of microearthquakes along creeping faults, *Nature* **400**, 635–641.
- Schaff, D. P. (2001). 4D high resolution seismology: repeating events and large scale relocation, *Ph.D. Thesis*, Stanford University, Stanford, California, 115 pp.
- 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**, no. B9, 2186, doi 10.1029/2001JB000633.

- Shearer, P. M. (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.
- Waldhauser, F. (2001). A computer program to compute double-difference hypocenter locations, *U.S. Geol. Surv. Open File Rept. 01-113*, 25 pp.
- 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., W. L. Ellsworth, and A. Cole (1999). Slip-parallel seismic lineations along the northern Hayward fault, California, *Geophys. Res. Lett.* **26**, 3525–3528.

Department of Geophysics  
Stanford University  
Stanford, California 94305-2215  
(D.S., G.H.R.B., E.Z., F.W., G.C.B.)

U.S. Geological Survey  
345 Middlefield Road, MS/977  
Menlo Park, California 94025-3591  
(W.L.E.)

Manuscript received 6 December 2002.