Southern California Hypocenter Relocation with Waveform Cross-Correlation, Part 2: Results Using Source-Specific Station Terms and Cluster Analysis

by Peter Shearer, Egill Hauksson, and Guoqing Lin

Abstract We obtain precise relative relocations for more than 340,000 southern California earthquakes between 1984 and 2002 by applying the source-specific station-term (SSST) method to existing P- and S-phase picks and a differential location method to about 208,000 events within similar-event clusters identified with waveform cross-correlation. The entire catalog is first relocated by using existing phase picks, a reference 1D velocity model, and the SSST method of Richards-Dinger and Shearer (2000). We also perform separate relocations of Imperial Valley events by using a velocity model more suited to this region. Next, we apply cluster analysis to the waveform cross-correlation output to identify similar-event clusters. We relocate earthquakes within each similar-event cluster by using the differential times alone, keeping the cluster centroid fixed to its initial SSST location. We estimate standard errors for the relative locations from the internal consistency of differential locations between individual event pairs; these errors are often as small as tens of meters. In many cases the relocated events within each similar-event cluster align in planar features suggestive of faults. We observe a surprising number of such faults at small scales that strike nearly perpendicular to the main seismicity trends. In general, the fine-scale details of the seismicity reveal a great deal of structural complexity in southern California fault systems.

Introduction

Waveform cross-correlation is an increasingly important tool for characterizing event similarity, improving earthquake locations, and studying source properties (Nakamura, 1978; Poupinet et al., 1984; Got et al., 1994; Dodge et al., 1995; Nadeau et al., 1995; Gillard et al., 1996; Rubin et al., 1999; Waldhauser et al., 1999). However, it is not yet used routinely for network data because of its greater computational requirements compared with standard processing based on phase picks. For example, even relatively small clusters (thousands of events) require that millions of cross-correlation functions be computed and that large parts of the waveform archive be available online. However, with modern computers these requirements are increasingly tractable and larger and larger problems may be addressed.

Here we build on our previous work with waveform cross-correlation location in southern California (Shearer, 1997, 1998; Astiz et al., 2000; Astiz and Shearer, 2000; Shearer, 2002; Shearer et al., 2003) to process and relocate the complete southern California earthquake catalog from 1984 to 2002. Although many of our methods are similar to previous work, we have made some changes and developed several new algorithms to handle the larger number of events in the complete catalog. In general, our new event locations are much more tightly clustered than previous catalogs, often aligning in linear and/or planar structures suggestive of faults. These newly identified features should help to unravel the complicated fault geometries and tectonics in southern California.

This is a companion paper to Hauksson and Shearer (2005), which describes the waveform cross-correlation calculation in greater detail and presents results of applying the double-difference location method (Waldhauser and Ellsworth, 2000, 2002).

Source-Specific Station Term (SSST) Locations

Before using the waveform cross-correlation times, we first relocate the catalog using a 1D velocity model and existing P- and S-phase-pick information. These results provide the mean locations of the similar-event clusters that we will analyze later, as well as background locations for those events that are not part of these clusters. We apply the SSST algorithm of Richards-Dinger and Shearer (2000) separately to the events within each of the polygons described previ-
ously. This method uses a grid-search, L1-norm approach to locate the events, iterating to generate custom station terms that are specific to each source region. This greatly improves the relative location accuracy of nearby events by accounting for the biasing effects of travel-time anomalies caused by 3D velocity structure.

Our current algorithm is modified slightly from Richards-Dinger and Shearer as follows. (1) Travel times between the stations and candidate event locations are extracted from simple distance/depth tables as needed rather than precomputed and stored on disk for every station to a 3D grid of possible source locations as in Richards-Dinger and Shearer. This simplifies the code and storage requirements at the cost of somewhat increased computation time. (2) During the iterations for the station terms, we gradually reduce \( n \), the number of adjacent events that we smooth over, from 1000 to 20. This seems to give superior results compared with when we begin the calculation with small values of \( n \). (3) We use a velocity-depth model (Fig. 1a) with somewhat lower velocities near the surface and a steeper lower crustal velocity gradient than the model used in Richards-Dinger and Shearer (2000), which is a smoothed version of the standard Hadley and Kanamori (1977) model for southern California. This reduces the number of events that locate at depths between 0 and 2 km where it is unlikely that large numbers of earthquakes are occurring. A histogram of the depth distribution of the SSST locations is shown in Figure 2.

A 1D velocity model cannot account for lateral variations in structure across southern California, so the accuracy of our absolute locations is limited, in particular, in depth (which is more difficult to constrain than horizontal position). Biases in depth are likely to be particularly severe in regions where the local velocity structure differs substantially from the 1D model, such as the Salton Trough and the offshore Continental Borderland region. In these areas, much better results can be obtained by using custom velocity models (Astiz and Shearer, 2000; Shearer, 2002). To obtain a more accurate set of hypocenters for the Imperial Valley region, we performed separate relocations for events in this area with a model (see Fig. 1b) derived from refraction seismic experiments (Fuig et al., 1984; Shearer, 2002).

Identifying Similar-Event Clusters

The next step in our processing is to use the waveform cross-correlation results to identify clusters of similar events (see Hauksson and Shearer [2005] for a description of the cross-correlation procedure). The output of the cross-correlation calculation contains information about the similarity of selected pairs of events. A range of event similarity exists, and there is some subjectivity in deciding whether an event pair is similar enough to yield good differential time data. After limited experimentation, we adopted the criteria that the event pair must have 8 or more \( P \) or \( S \) measurements with correlation coefficients above 0.65 for stations within 80 km of the events. We adopted the range cutoff because seismograms from distant stations are generally lower in frequency and tend to give high correlations even when the events are not very similar because of the limited bandwidth of the records. There are often several measurements from
different components of the same station. We remove this redundancy before applying our selection criteria by favoring \( P \) measurements from the vertical component and \( S \) measurements from the horizontal components, and then selecting the measurement with the highest correlation.

For those event pairs that exceed our similarity cutoff, we compute the mean correlation coefficient of the individual \( P \) and \( S \) values to use as an overall measure of the similarity of the pair. Next, we apply a cluster analysis approach (Hartigan, 1975) to identify groups of events that are correlated with each other. Because similarity information is not available for every possible event pair, some modification to standard cluster analysis algorithms was necessary to achieve a suitable method. We first sort the event-pair correlation coefficients. Beginning with the most correlated event pair, we group the events into a single-event cluster. This continues for decreasing values of event-pair similarity. When a link is found between an event pair, on which at least one of the events is already in a similar-event group, we check to see whether additional links exist between the two existing groups. The two clusters are combined into a single cluster only if a specified fraction of the possible links have measurements that indicate similar events. This avoids the possibility of a single false link between two clusters causing them to collapse into a single cluster. In practice, we found that requiring 1% of the possible links to indicate event-pair similarity was sufficient to prevent significant false grouping of clusters.

Recently, Du et al. (2004) examined some of the effects of different cross-correlation methods and of varying the cross-correlation cutoff on relocation results obtained with the double-difference algorithm (Waldhauser and Ellsworth, 2000). Our choices in defining similar-event clusters are specific to our particular dataset and processing method and will not necessarily be appropriate for other studies. In general, the degree of waveform cross-correlation depends on the window length and the frequency content of the data (as filtered before cross-correlation). After some experimentation, we applied relatively low cross-correlation cutoffs to maximize the number of cross-correlation measurements that we used, relying on the robust L1-norm differential location method of Shearer (2002) to minimize the effects of spurious correlations. For similar reasons, we required only 1% of possible links to exist in defining similar-event clusters because we wanted to maximize the number of connections between event pairs and found that the 1% requirement was sufficient to prevent the "collapsing" of large numbers of distributed events into single clusters. Again this cutoff is likely specific to our dataset, which computes cross-correlations only between every event and 100 neighboring events, so results are not available for all possible event pairs. Given the size of our dataset, a full exploration of the effects of various applied cutoffs (correlation threshold, required number of stations, maximum distance, cluster link fraction, etc.) is beyond the scope of this article.

Those events within similar-event clusters are then relocated by using the waveform cross-correlation data alone, using all event pairs within the cluster that are defined as similar by using the criteria discussed previously (eight or more correlation coefficients above 0.65 for stations within 80 km of the event). Once an event pair is defined as similar, however, we use all differential times in which the correlation coefficient is 0.6 or higher. We solve for a single best-fitting set of locations for all the events in the cluster with the method discussed in Shearer (2002) and Shearer et al. (2003). Estimated standard errors in the relative locations are obtained from the internal consistency of the differential locations among the different event pairs. The mean location of each cluster remains fixed at its original SSST position as the differential times have little sensitivity to absolute location. Clusters are only relocated if they contain five or more events. Of the events, 207,628 (61% of the original catalog) are in similar event clusters and are relocated. Events that are not contained in similar-event clusters are not relocated and retain their SSST locations, based on phase-pick data alone.

Comparisons between Location Methods

A 1984 to 1985 swarm of events near Vallecito Valley contains a sharp northeast-striking seismicity alignment (named the Sawtooth Ridge fault by Magistrale and Rockwell, 1996) that provides a good test of differences in location methods. In Figures 3 and 4 we compare locations of about 2400 events obtained from (a) the Southern California Seismic Network (SCSN) catalog, (b) the SSST results of
Figure 3. Computed epicenters of seismicity near Vallecitos Valley, comparing: (a) SCSN catalog locations, (b) SSST locations from Richards-Dinger and Shearer (2000), (c) double-difference locations from Hauksson and Shearer (2005), and (d) the cluster analysis locations discussed in this article. (a) and (b) are from phase-pick data alone; (c) and (d) use differential times obtained using waveform cross-correlation. The black dots in (d) are events within similar event clusters; the gray dots are uncorrelated events that are located using phase-pick data alone. The line AB shows the cross section for Figure 4. This feature was named the Sawtooth Ridge fault by Magistrale and Rockwell (1996).

Richards-Dinger and Shearer (2000), (c) the double-difference method (Waldhauser and Ellsworth, 2000, 2002, as applied by Hauksson and Shearer, 2005), and (d) the cluster analysis method used here. Note that methods a and b use phase-pick data alone; methods c and d use waveform cross-correlation results. The catalog locations show the most scatter, particularly in depth, where many locations are fixed to 6 km. The SSST method sharpens the image somewhat, but the most dramatic improvements in the locations are obtained with the waveform cross-correlation results. The double-difference and cluster analysis hypocenters agree reasonably well. Some of the variations in absolute locations are a result of differences in the velocity models, with the double-difference results obtained by using as a starting
Figure 4. Cross sections of seismicity along the profile AB (shown in Fig. 3) for seismicity near Vallecitos Valley, comparing: (a) SCSN catalog locations, (b) SSST locations from Richards-Dinger and Shearer (2000), (c) double-difference locations from Hauksson and Shearer (2005), and (d) the cluster analysis locations discussed in this article. See Figure 3 for additional details.

point the Hauksson (2000) catalog based on a 3D velocity model. Notice that the locations presented here have fewer events below 2 km than the Richards-Dinger and Shearer (2000) catalog; this is primarily a result of using a different 1D velocity model (see previously).

The locations plotted in Figures 3d and 4d are of two types. The most accurate relative locations are obtained for the similar-event clusters and are shown as black dots. Forty-five separate similar-event clusters are shown in the plot. Events that do not cross-correlate well enough with other events to be part of a similar-event cluster retain their original SSST locations and are plotted as light-gray dots. Estimated uncertainties for the relative location accuracy within individual similar-event clusters vary considerably among
the different clusters, but typically have estimated horizontal standard errors less than 50 to 100 m and estimated vertical standard errors of less than 100 to 200 m. The best constrained events have estimated standard location errors of tens of meters.

The Complete Catalog

Locations were first obtained separately for each of the five different polygon-shaped regions (see figure 1 in Hauksson and Shearer, 2005). However, we do not include results from polygon 1 in our final catalog because it is on the edge of the southern California network and we did not obtain good results for these events. Because of the slight overlap among the polygons, some events are located in more than one polygon. For the final catalog, one of the possible locations was randomly selected. This catalog, SHLK_1.0, is available with documentation regarding its format at the Southern California Earthquake Center (SCEC) web site (www.data.scec.org) and also via anonymous ftp to mahi.ucsd.edu in the /pub/Shearer/SHLK directory. A separate catalog, SHLK_IMP_1.0, with more accurate locations for the Imperial Valley region is also available (these events are included in SHLK_1.0 but are located less accurately because of the inadequacies of the 1D model used for all of southern California when applied to the anomalous seismic velocities in the Salton Trough).

The relocated earthquake catalog includes events located with phase-pick data alone (colored by year of occurrence in Fig. 5) and events within similar-event clusters, which have been relocated using the waveform cross-correlation data (shown in black in Fig. 5). The greatest density of similar events is found in the northern part of this plot, for example, in the Ridgecrest area. A striking feature of the new locations is the number of northeast-trending features that are nearly orthogonal to the northwest trend of the many of the major faults, such as the San Andreas and San Jacinto. A good example of this is the Vallecito swarm plotted in Fig. 3. Other prominent examples can be found in the Salton Trough. Figure 6 plots events in this region located by using the special velocity model for the Imperial Valley. In addition to the well-known northeast trend of the 1987 Elmore Ranch earthquake, much of the seismicity within the Brawley Seismic Zone (BSZ) is organized along a series of northeast-striking faults that are nearly orthogonal to the overall trend of the BSZ (Johnson and Hill, 1982; Hudnut et al., 1989b). The northern part of the BSZ, however, appears to have northwest-trending faults along its northeast boundary. As suggested by Magistrale (2004), these may be a continuation of the San Andreas Fault (SAF). There is some seismicity northeast of the Salton Sea near the surface expression of the SAF. Although these events are aligned parallel to the SAF, they are displaced to the northeast of the fault trace (Magistrale, 2004). Earthquakes near the Imperial Fault, just north of the United States/Mexico border form a series of streaks parallel to the fault at about 10 km depth.

These features are discussed in more detail in Shearer (2002).

In general, these results agree with previous cross-correlation studies in specific regions (Shearer, 1997, 1998; Astiz et al., 2000; Shearer et al., 2003) but there are often differences in absolute depth related to the velocity model used in the SSST locations. For example, the Whittier Narrows aftershocks are located at ~13 km depth, about 3 km shallower than in Shearer (1997) but close to the revised depths obtained by Shaw and Shearer (1999). The Oak Ridge cluster is located at 13 km depth rather than the 18 km obtained by Shearer (1998). The revised absolute location for the Oak Ridge cluster affects the dip direction of the seismicity plane resolved by the waveform cross-correlation; our new results show the dip direction to be just west of north rather than northeast as resolved by Shearer (1998). Because the true depth for this cluster is unknown, it is not clear which study is more accurate; the uncertainty in the dip direction is probably 45 degrees or more.

The tendency for earthquakes to be part of similar-event clusters varies across southern California, as illustrated in Figure 7, which plots the fraction of events that are contained in the similar-event clusters as a function of epicenter position. The highest percentages of similar events are found in the northernmost part of the network in the Ridgecrest and Coso areas, where the number of similar events typically exceeds 80%. Artificial seismicity, such as the quarry blasts seen in the Mojave block, very rarely forms similar-event clusters. This may be a result of ripple firing and other source complexity, as well as strong heterogeneity in the near surface, which results in nonrepeatable waveforms. The Landers and Hector Mine aftershock sequences yield about 50% to 70% similar events. The Northridge aftershocks contain about 70% to 80% similar events.

Of course, the fraction of similar events will depend on the details of the waveform similarity required to define similar-event pairs. Because our focus is on improving event locations, we include events that cross-correlate well enough to produce reliable differential times even if the waveforms are not wiggle-for-wiggle identical. The similar-event clusters defined here are not repeating, nearly identical events, such as those seen at Parkfield, California, by Nadeau et al. (1995). Because of our requirement that at least eight similar waveforms exist from stations within 80 km of the epicenter, events at the edge or outside of the network will likely yield fewer similar events than events in the middle of the network. For example, there are similar events in the offshore Oceanside aftershock sequence (Astiz and Shearer, 2000) located near 33°N, 117.8°E, but these are not included as similar in the present study because of the 80-km station-distance cutoff requirement.

Example: Scodie Lineament

As an example of the resolution of our seismicity catalog, Figure 8 plots events in the southern Sierra Nevada
region. A diffuse band of seismicity, termed the Scodie lineament by Bawden et al. (1999), exists between the 1952 $M_{7.3}$ earthquake on the White Wolf Fault and the 1946 $M_{6.1}$ Walker Pass earthquake (located near the Scodie Mountains in Fig. 8). Aftershocks of the Walker Pass event are still observed today and form a wedge-shaped seismicity pattern near the Scodie Mountains. Bawden et al. (1999) speculate that the Scodie lineament represents a newly forming, incipient strike-slip fault. Our locations define the sharpness of the seismicity features in this region with clarity similar to the relocations of Bawden et al., who used data from both northern and southern California stations, a site-specific velocity model, and a joint hypocenter location method. In some cases, the relocated similar-event clusters provide de-
Figure 6. Seismicity in the Salton Trough located with a custom velocity model for the Imperial Valley. Notice the northeast-striking faults in the Brawley Seismic Zone.
Figure 7. The fraction of events that are contained in the similar-event clusters that were computed in this study from waveform cross-correlation. The numbers are plotted at the center of 15 by 15 km cells, with 0 indicating from 0% to 10% similar events, 1 indicating from 10% to 20% similar events, etc. Cells with fewer than 20 total events are not plotted.

tail not visible in the Bawden et al. study. For example, the seismicity at the northern end of the White Wolf Fault (near 35.3°N, −118.55°E) is organized into a series of northwest-trending faults that are nearly orthogonal to the northeast trend of the Scodie lineament. Seismicity on the Garlock Fault plotted in Figure 8 forms a linear trend at about 10 km depth displaced 2 to 3 km to the north of the surface expression of the fault. Seismicity in the southern Sierra Nevada between the White Wolf Fault and the Scodie Mountains is diffuse, but those similar-event clusters that are clearly imaged tend to trend north–south. These north and northwest-trending seismicity alignments within the earthquakes along the Scodie lineament suggest complexities in the tectonics that cannot be explained with a single north-east-trending fault.

Discussion

Several studies have identified streaks of microearthquakes along active fault surfaces in northern California (Rubin et al., 1999; Waldhauser et al., 1999; Rubin and Gilillard, 2000; Waldhauser and Ellsworth, 2000). We do not observe comparable features in southern California, with the possible exception of seismicity streaks along the Imperial Fault (Shearer, 2002). The Imperial Fault streaks, however, are not confined to a single fault plane but rather appear to indicate multiple faults at about 9 km depth. The along-fault streaks in northern California are observed on or near creeping sections of major faults; the absence of comparable regions of deep fault creep in southern California may explain the apparent absence of seismicity streaks. There are none-
thereby many similar-event clusters in southern California that can be used to image small-scale fault structures. In general, these suggest a large degree of fault complexity across the region. Fault orientations inferred from similar-event clusters will be useful in unraveling the tectonics and computing stress orientations, in particular, when the results are combined with focal-mechanism analyses (Shearer et al., 2003).

Our results demonstrate that it is now possible to apply advanced earthquake location methods, including waveform cross-correlation and differential location algorithms, to large local earthquake catalogs (Got and Okubo, 2003; Wolfe et al., 2003). A long-term goal of our research is to begin implementing these methods into standard network practice (Got et al., 2002). A difficulty with adding such methods to real-time processing is that in these approaches earthquake locations are not computed independently of other events. Thus, in principle the entire analysis needs to be redone whenever a new event is added to the catalog and the locations of older events in the catalog could change. As a practical matter, it would make sense to redo the entire catalog only at fixed periods, such as once per year. Between these complete relocation efforts, waveforms from new events would be cross-correlated to older events and their locations computed with respect to the older events, whose locations would remain fixed. This approach would probably require modification, however, if a new cluster of seismicity develops in an area with little or no previous seismicity. In this case, the relative relocation methods used here would need to be applied to all the events in the cluster simultaneously.

As computer capabilities continue to grow, it will be
possible to improve our methods by increasing the fraction of event pairs on which we perform waveform cross-correlation. Ultimately the differential times could be integrated into inversions for 3D velocity structure, such as the double-difference tomography described by Zhang and Thurber (2003), which would in turn improve the absolute event location accuracy. As demonstrated by Rubin (2002), the differential times also provide precise calibration information for detecting station and network timing discrepancies. Finally, the online waveform database that we have created will be useful for other studies, such as amplitude and spectral analyses of source properties and crustal attenuation structure.

Conclusions

Precise relative locations of more than 340,000 southern California earthquakes provide new insights into fault structures and earthquake clustering. In general, the clustering of seismicity tightens around major mapped late Quaternary faults. In addition, we can identify numerous small seismicity alignments that form high angles to mapped faults. These alignments suggest that off-fault deformation is often accommodated by high-angle cross-faults, thus implying possible small-scale block rotations (Nicholson et al., 1986; Hudnut et al., 1989a). On average 61% of the earthquakes are in similar event clusters but there are significant differences in similar event density across southern California, with the greatest fraction of similar events found in the Ridgecrest-Coso area.

Acknowledgments

We thank the personnel of the USGS/Caltech Southern California Seismic Network who pick and archive the seismograms and the Southern California Earthquake Center (SCEC) for distributing the data; we thank Vicki Appel, who facilitated access to the database. Cliff Thurber and an anonymous reviewer provided constructive comments. Funding for this research was provided by NEHRP/USGS Grant 03HQPA0001. This research was also supported by the Southern California Earthquake Center, which is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. SCEC contribution no. 784. Contribution 9106 of Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena.

References


Shearer, P. M. (1997). Improving local earthquake locations using the L1 norm and waveform cross-correlation: a application to the Whittier Nar-


Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography
University of California at San Diego
La Jolla, California 92037-0225
pshearer@ucsd.edu
(P.S., G.L.)

Seismological Laboratory
California Institute of Technology
Pasadena, California 91125
hauksson@gps.caltech.edu
(E.H.)