

# Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005

Guoqing Lin,<sup>1,2</sup> Peter M. Shearer,<sup>1</sup> and Egill Hauksson<sup>3</sup>

Received 14 February 2007; revised 27 August 2007; accepted 13 September 2007; published 29 December 2007.

[1] We compute high-precision earthquake locations using southern California pick and waveform data from 1981 to 2005. Our latest results are significantly improved compared to our previous catalog by the following: (1) We locate events with respect to a new crustal  $P$  and  $S$  velocity model using three-dimensional ray tracing, (2) we examine six more years of waveform data and compute cross-correlation results for many more pairs than our last analysis, and (3) we compute locations within similar event clusters using a new method that applies a robust fitting method to obtain the best locations satisfying all the differential time constraints from the waveform cross correlation.

These results build on the relocated catalogs of Hauksson and Shearer (2005) and Shearer et al. (2005) and provide additional insight regarding the fine-scale fault structure in southern California and the relationship between the San Andreas Fault (SAF) and nearby seismicity. In particular, we present results for two regions in which the seismicity near the southern SAF seems to align on dipping faults.

**Citation:** Lin, G., P. M. Shearer, and E. Hauksson (2007), Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005, *J. Geophys. Res.*, *112*, B12309, doi:10.1029/2007JB004986.

## 1. Introduction

[2] Earthquake locations are fundamental parameters for studies of earthquake physics, fault orientation, and Earth's deformation. Studies of earthquake location improvements have been an important branch in seismology for the past few decades. To improve absolute location accuracy, we need knowledge of Earth's three dimensional velocity structure. For local earthquakes, this is usually done by simultaneously solving for a three-dimensional (3-D) velocity model and earthquake locations [e.g., *Thurber*, 1983, 1992; *Thurber and Eberhart-Phillips*, 1999; *Zhang and Thurber*, 2003].

[3] Recently, some techniques have been presented that are able to improve significantly the relative location accuracy among nearby events, even when the arrival times are biased by the effects of three-dimensional velocity structure [e.g., *Richards-Dinger and Shearer*, 2000; *Waldhauser and Ellsworth*, 2000; *Nicholson et al.*, 2004; *Lin and Shearer*, 2005]. Improvements in relative location accuracy obtained

using these methods often produce a dramatic sharpening of seismicity patterns. With the development of modern computers, waveform cross correlation has also been an increasingly important tool for improving relative earthquake locations because of the great accuracy of differential times [*Nakamura*, 1978; *Got et al.*, 1994; *Dodge et al.*, 1995; *Nadeau et al.*, 1995; *Gillard et al.*, 1996; *Rubin et al.*, 1999; *Waldhauser et al.*, 1999; *Moriya et al.*, 2003; *Hauksson and Shearer*, 2005; *Shearer et al.*, 2005].

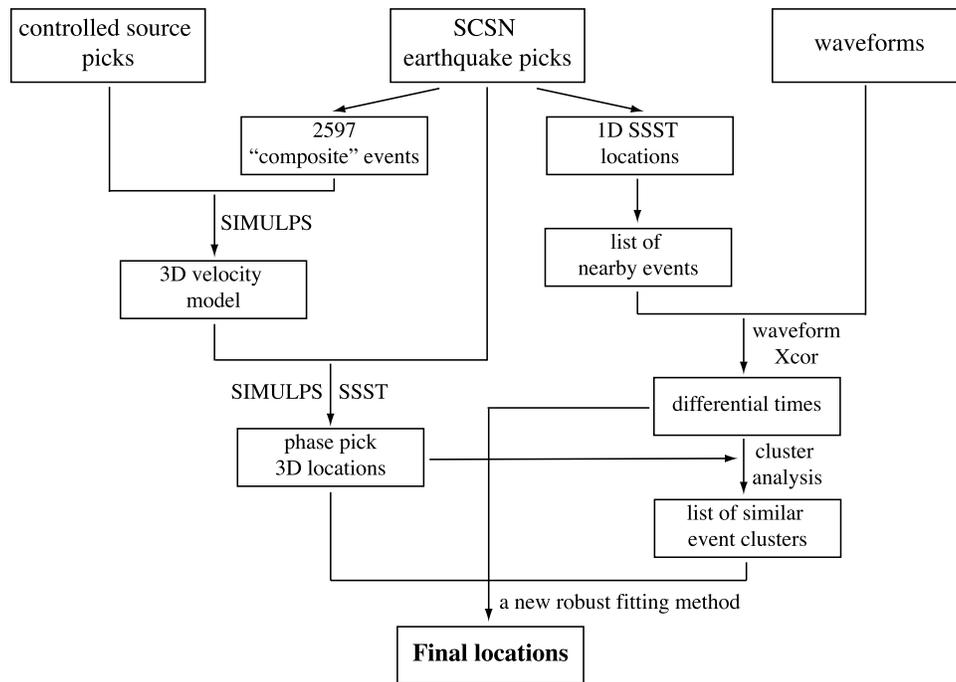
[4] Improved earthquake locations help to improve resolution of fault structures and characterize the spatial and temporal characteristics of seismicity. High-resolution event catalogs in southern California have recently been used to study the decay of aftershock density with distance [*Felzer and Brodsky*, 2006], explore the spatial relationship between aftershocks and main shock rupture planes [*Liu et al.*, 2003; *Powers and Jordan*, 2005], analyze the fractal dimension of seismicity [*Kagan*, 2006], and assess the mechanisms driving seismic swarms [*Lohman and McGuire*, 2007].

[5] In this study, 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; *Hauksson and Shearer*, 2005; *Shearer et al.*, 2005] to process and relocate the complete southern California earthquake catalog from 1981 to 2005. This results in six more years of data than the 1984–2002 SHLK (Shearer, Hauksson, Lin and Kilb) catalog [*Shearer et al.*, 2005]. Although many of our methods are similar to our prior work, we have made some

<sup>1</sup>Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA.

<sup>2</sup>Now at Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, Wisconsin, USA.

<sup>3</sup>Seismological Laboratory, California Institute of Technology, Pasadena, California, USA.



**Figure 1.** A work flowchart of our location procedures in this study. This paper is focused on the 1-D, 3-D SSST, and differential time locations.

changes and developed several new algorithms to handle the larger number of events in the complete catalog. We now locate the events using a new 3-D  $P$  and  $S$  crustal velocity model for southern California and a new robust least squares method to relocate events within similar event clusters using the waveform cross-correlation times. In addition, we include estimates of absolute and relative location errors. Our complete location procedure is outlined in the flowchart of Figure 1 and will be discussed in detail in sections 2–8. This paper is focused on the 1-D, 3-D source-specific station term (SSST), and differential time locations.

## 2. Locations From Phase Picks and a 3-D Velocity Model

[6] To obtain accurate absolute starting locations for waveform cross-correlation relocation, we use the new 3-D crustal  $P$  and  $S$  velocity models of *Lin et al.* [2007] to relocate all seismicity in southern California from 1981 to 2005 while keeping the velocity model fixed. We use the 3-D ray tracing capability of the SIMULPS algorithm [*Thurber*, 1983, 1993; *Eberhart-Phillips*, 1990; *Evans et al.*, 1994] to compute locations for all events with respect to the tomography model while iteratively adjusting the pick times using a source-specific station term approach [*Richards-Dinger and Shearer*, 2000; *Lin and Shearer*, 2005] to improve the relative locations among nearby events.

### 2.1. Data Sets

[7] Our data for the initial event locations are the phase arrival times of  $P$  and  $S$  waves from 452,943 events, including local events, regional events and quarry blasts, recorded at the Southern California Seismic Network

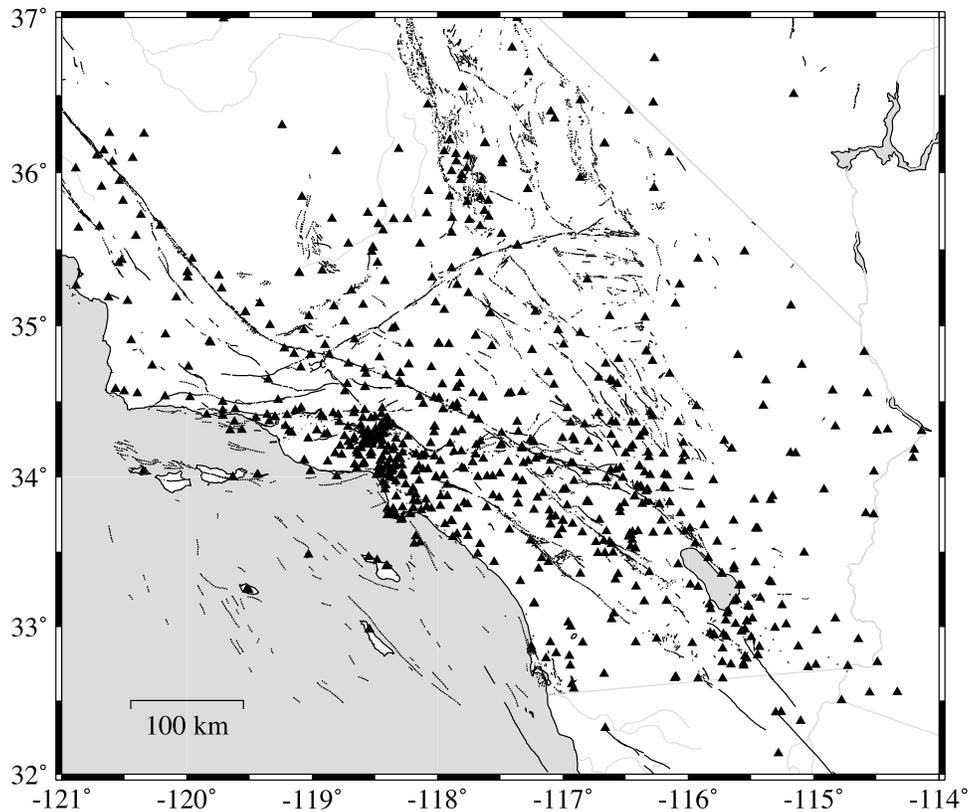
(SCSN) stations and picked by the network operators. Figure 2 shows the station locations in our study area. We require each event to have at least 5 observations ( $P$  and  $S$  picks) from stations within a 150 km distance cutoff. This results in about 430,000 events to be relocated using 3-D ray tracing. To refine the relative locations among closely spaced events, we combine the 3-D ray tracing with the source-specific station term relative location method.

### 2.2. Shrinking Box Source-Specific Station Term Method

[8] The source-specific station term (SSST) method improves relative event locations among nearby events using phase arrival times [*Richards-Dinger and Shearer*, 2000; *Lin and Shearer*, 2005]. This method attempts to correct for the systematic biases in arrival times caused by three-dimensional velocity variations without actually solving for the velocity structure itself. The station corrections are calculated for each source-receiver pair at a given station using the residuals from nearby events within a given distance cutoff separately for  $P$  and  $S$ , so the station correction varies as a function of source position. The station term part of the calculation is separate from the event location so the method can be applied using any desired location technique. The shrinking box SSST is an extension of the simple SSST in that it computes the SSST terms while continuously shrinking the cutoff distance between the first and final iteration. For more details, please refer to *Lin and Shearer* [2005].

### 2.3. Combination of 3-D Ray Tracing With SSST

[9] The crustal structure in the 3-D velocity model provides improved absolute hypocenter locations by correcting for the biasing effects of large-scale velocity varia-



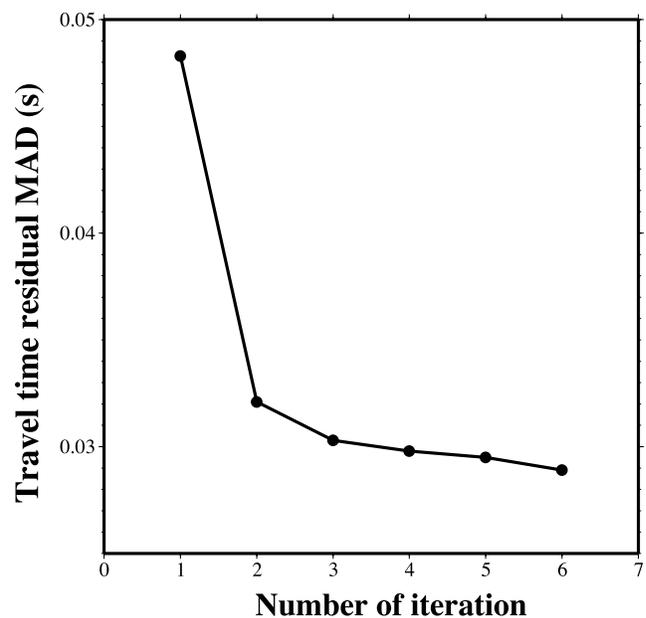
**Figure 2.** Locations of the 783 stations used in our study area.

tions. However, the tomography model is relatively smooth and cannot account for small-scale velocity structure that can also introduce bias and scatter in event locations. Thus to further refine the event locations, we combine the shrinking box SSST method and 3-D ray tracing in Thurber's SIMULPS computer algorithm. Our strategy is to first relocate all seismicity using the 3-D ray tracing, then compute SSSTs for each individual pick from the traveltime residuals of nearby events. Next, we subtract the SSST terms from the arrival time picks, and repeat the 3-D relocations with the new traveltime data. We perform 6 iterations of 3-D location and SSST computation. The distance cutoff for the station term calculation is reduced gradually during the iterations from 100 km to 10 km. We find that this approach converges quickly to a stable set of locations and station terms. Figure 3 shows the reduction of the traveltime residual median absolute deviation (MAD) from the 430,000 events with iteration number in our SSST calculation. The MAD of the residuals drops from 0.048 to 0.029 s. For comparison, the root-mean-square (RMS) residual decreases from 0.16 to 0.12 s. The event locations at this point represent our best estimates based on phase pick data alone without the further improvements that are possible using differential times from waveform cross correlation.

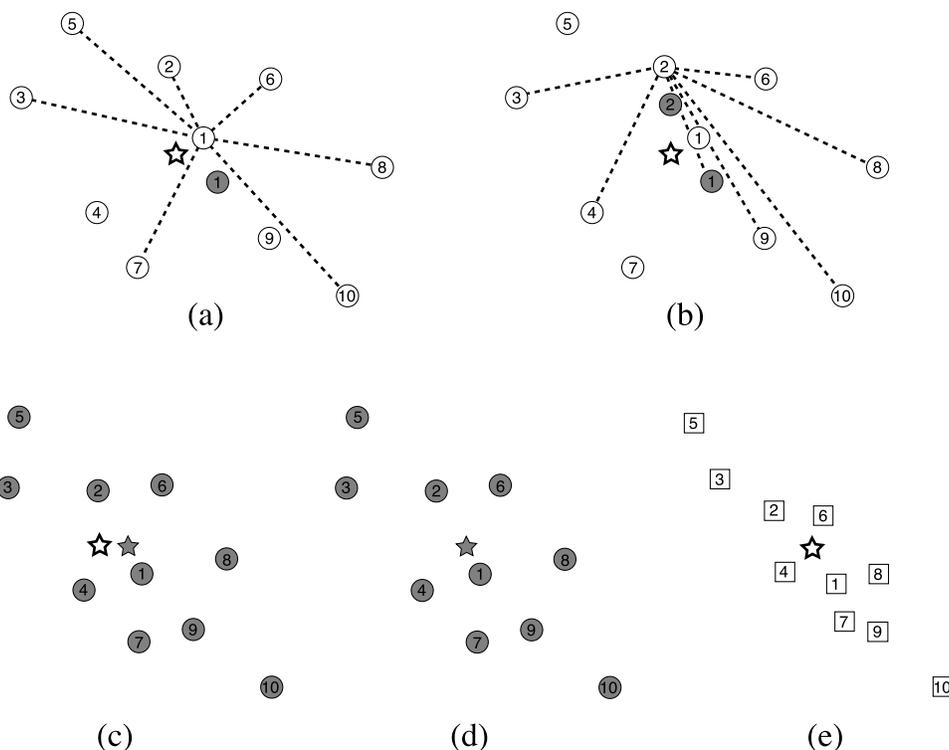
### 3. Waveform Cross Correlation

[10] The waveforms of nearby earthquakes recorded at the same station are often similar enough that waveform cross correlation can be used to obtain much more precise

differential times than can be picked on individual seismograms, in which case greatly improved relative locations among the events can be computed. The waveform cross-correlation process in this study is similar to that described by *Hauksson and Shearer [2005]* and *Shearer et al. [2005]*.



**Figure 3.** Reduction of traveltime residual MAD with iteration numbers of SSST calculations.



**Figure 4.** Cartoons showing our new differential time location method. (a) Starting locations (open circles) for the 10 events in a similar event cluster. The star is the centroid of the cluster. After fitting the differential times between event 1 and the linked events, event 1 is shifted to the gray circle. (b) Relocation of event 2 using the new location of event 1 if they are correlated with each other. (c) Distribution of the cluster after all the events are relocated shown by the gray circles. Note that the new centroid of the cluster (the gray star) is different than the starting centroid. (d) Shifted cluster location centered at the starting centroid. (e) Final locations for this cluster after repeat of Figures 4a–4d for a few iterations.

### 3.1. Waveform Data Processing

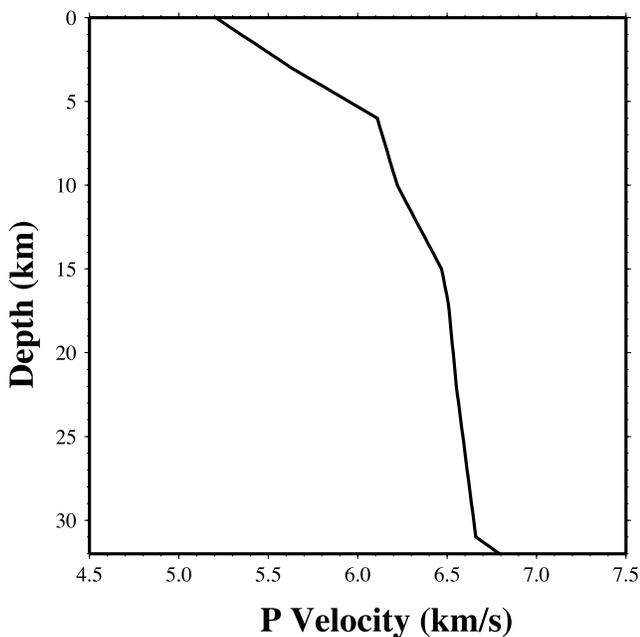
[11] We obtain waveform data from the SCSN, as archived at the Southern California Earthquake Data Center (SCEDC) for all available local events, regional events and quarry blasts in southern California from 1981 to 2005. We use the Seismic Transfer Program (STP) ([www.data.sceec.org/STP/stp.html](http://www.data.sceec.org/STP/stp.html)) to extract the waveforms in Seismic Analysis Code (SAC) format ([www.llnl.gov/sac/](http://www.llnl.gov/sac/)). We obtain all components (e.g., vertical, east, west) and channels (short-period, broadband, etc.) contained in the database archive. Our first step is to trim the seismograms to 60 sec, starting 10 sec before the theoretical  $P$  arrival time. We store the resulting time series using the event-based GFS format (G. Masters, personal communication, 2006) within a year/month directory structure. Following these steps, the GFS files consume about 626 Gb on an online Redundant Array of Independent Disks (RAID) system that provides rapid and random access to the data. We then resample the data to a uniform 100-Hz sample rate (using a spline interpolation method) and apply a bandpass filter between 1 and 10 Hz.

### 3.2. Waveform Cross-Correlation Calculation

[12] It is computationally infeasible to cross correlate every event pair for all 450,000 events because the size of the problem scales as  $n(n - 1)/2$ . Thus we restrict the

calculation to event pairs separated by less than 2 km. However, to ensure a significant number of pairs even in regions of sparse seismicity, if the number of events within 2 km of an event is less than 100, we add additional events defined using a Delaunay tessellation [Richards-Dinger and Shearer, 2000] of our catalog until we have at least 100 neighboring events. 62% of the total events required the use of the Delaunay tessellation. We define the neighboring events using an event catalog based on 1-D locations using the shrinking box SSST method (at this point we had not yet computed the locations based on the 3-D model). Although the absolute location accuracy of this initial catalog is limited by the use of a 1-D model (the model used by Shearer *et al.* [2005]), the relative location accuracy is sufficient for us to use these locations to identify similar event pairs. In total we compute cross-correlation functions for all available station and components for over 94 million event pairs, about 7 times more pairs than we computed previously for the SHLK catalog [Shearer *et al.*, 2005].

[13] We compute the cross-correlation functions separately for  $P$  and  $S$  waves, applying symmetric time shifts of up to  $\pm 1.5$  s, using a spline interpolation method to achieve a nominal timing precision of 0.001 s (1 ms). If catalog picks are available, we use a 1.5 s window around  $P$  and a 2.5 s window around  $S$ . If picks are not available, we estimate arrival times from the earthquake location and a simple 1-D



**Figure 5.**  $P$  velocity model used for calculating theoretical traveltimes during cross-correlation relocation.

model, using a 2-s  $P$  window and a 3-s  $S$  window. Our pick windows are designed to avoid including part of the  $S$  wave in the  $P$  window or part of the  $P$  wave in the  $S$  window.

[14] Information is saved only for the 31 million event pairs with an average waveform correlation coefficient of 0.45 or greater and with at least 10 individual differential times with correlation coefficients of 0.6 or greater. The resulting correlation coefficients and time shifts contain information regarding the similarity of events and their relative locations. In our previous study [Hauksson and Shearer, 2005; Shearer et al., 2005], we used roughly the same criteria and obtained correlation information for about 3 million pairs of events. The tenfold increase in usable pairs reflects the greater number of pairs computed (six more years of data and more pairs per target event) and an increased yield of correlated pairs (obtained because a greater fraction of the pairs are separated by distances of less than 2 km). A variety of different methods can be applied to analyze cross-correlation data. In this study, we apply a cluster analysis method similar to that described by Shearer et al. [2005].

#### 4. Similar-Event Cluster Analysis

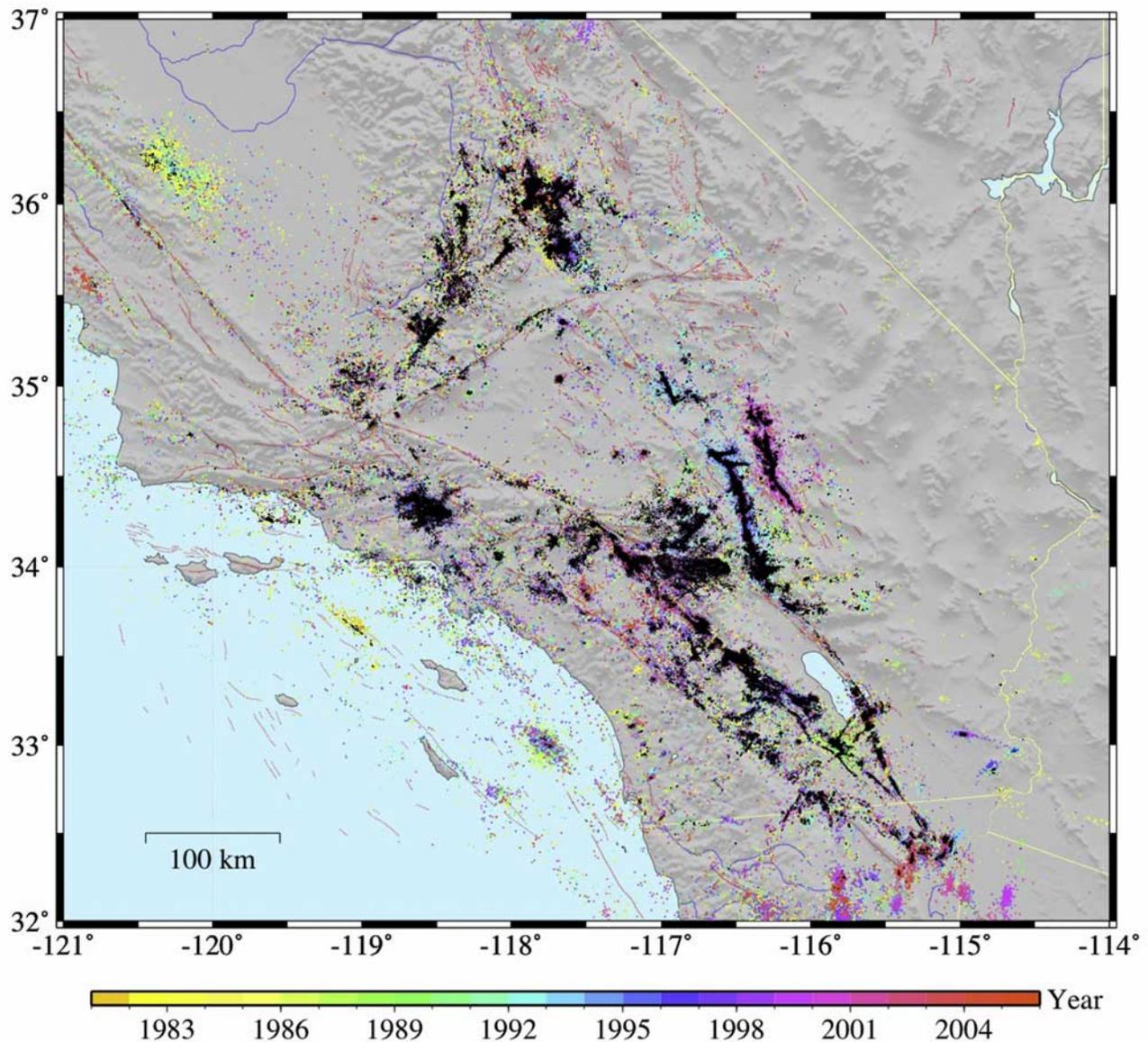
[15] The next step in our processing is to use the waveform cross-correlation results to identify clusters of similar events. The output of the cross-correlation calculation contains information about the similarity of selected pairs of events. On the basis of our experience, we adopted the criteria 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. 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 (the specifics of which are described by Shearer et al. [2005]) to identify groups of events that are correlated with one other. Then we use the waveform cross-correlation times to relocate the 323,000 events in 3676 similar event clusters with more than 5 events.

#### 5. Differential Time Relocation Method

[16] We use the differential times from the waveform cross correlation to relocate events within each similar event cluster in order to further improve the relative event locations. Here we describe a new method for computing these locations, which has some advantages compared to the technique we applied previously for the SHLK catalog [Shearer et al., 2005]. The location algorithm for the SHLK catalog performed differential locations for each correlated event pair in the similar event cluster and then reconciled all the differential locations into a single best fitting set of locations. Although this two-stage approach seems to work well in practice, the final locations are only indirectly determined by the original differential times. Our new method solves for the location for each event directly from all the differential times between this target event and the linked events in the cluster and repeats the procedure for a few iterations until the method converges to a stable set of locations. Because there is more differential time information for each event in the new method, the location is more robust. In addition, the algorithm for the SHLK catalog was based on a least squares method, while the new LSH (Lin, Shearer, Hauksson) catalog uses a robust least squares estimate, which is less affected by outliers in the differential times.

[17] The cartoons in Figure 4 show how our method works. In this example, there are 10 events, shown by the blank circles, in a similar event cluster centered at the star in Figure 4a. Each event is linked with some, but not all, of the other events in the cluster by differential times. The event number is sorted chronologically. We start with event 1, which is linked with 7 other events in this example. We apply a grid search algorithm to find the best location of event 1 (shown by the gray circle) that minimizes the robust least squares [Lin and Shearer, 2007] of the differential time residuals while keeping the locations of the 7 neighboring events fixed. The new fitting method measures distance using the  $L_2$  norm for data misfits below some specified value,  $d_{\max}$  (which in general will depend on the observations), and the  $L_1$  norm for larger values. This hybrid  $l^1 - l^2$  error measure was proposed by Huber [1973] and is relatively insensitive to outliers in data, so we term it the “robust least squares” method. This robustness is important because we want to use as many differential times as possible, even at the risk of including some falsely correlated waveforms that produce large residuals. In this study, we used  $d_{\max} = 0.1$  s. We relocate other events in the cluster in the same way except that we use the updated locations of the neighbors (see the relocation of event 2 in Figure 4b).

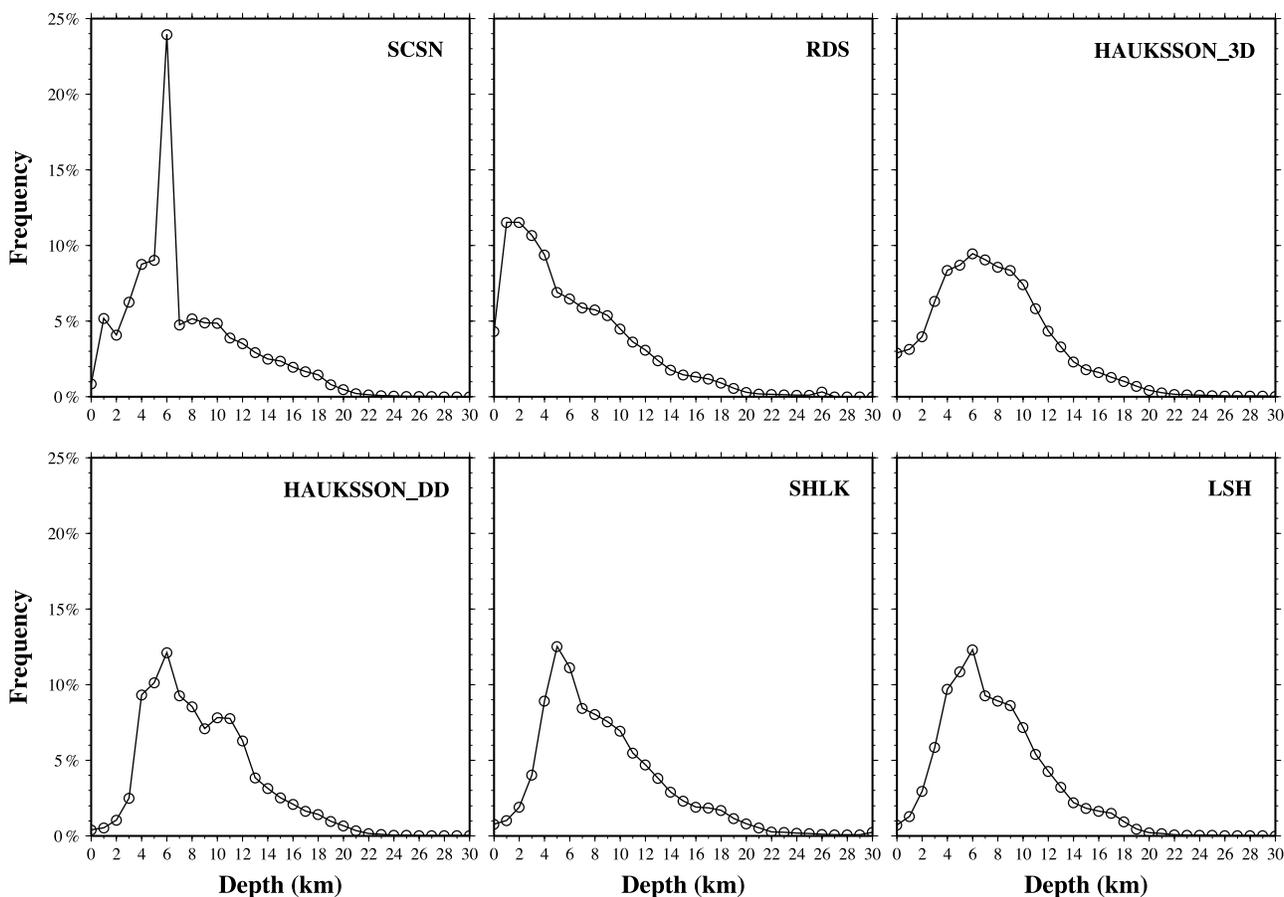


**Figure 6.** Map view of the relocated seismicity from 1981 to 2005 in our study. Black dots show similar event clusters relocated using cross-correlation data. About 25% of events do not correlate and are plotted in color by year at their 3-D locations or 1-D locations.

After one iteration, all the event locations in the cluster are updated to the gray circles in Figure 4c. Usually the centroid of the relocated cluster (the gray star) is slightly different than the initial centroid (the blank star). To stabilize the inversion and because the differential times are only very weakly sensitive to the absolute cluster location, after each iteration we shift the entire cluster so that the new centroid of the cluster is the same as the centroid of the starting locations, shown in Figure 4d. Thus the absolute cluster locations remain constrained by the locations based on the  $P$  and  $S$  picks and computed from ray tracing through the 3-D model. We repeat this process with the updated locations for a few iterations and generally observe rapid convergence to a stable set of locations (Figure 4e) that does not depend upon the initial event ordering. Using this

method, we separately relocate each of the 3676 similar event clusters in southern California.

[18] The relocations within each similar event cluster are performed entirely using the differential times obtained by waveform cross correlation; that is, we do not use any of the original picks at this stage. The pick information is used only to provide the initial locations that determine the centroid of each cluster. Thus we do not attempt to compute pick corrections, as described by *Aster and Rowe* [2000] and *Rowe et al.* [2002]. However, at least some of the effect of pick bias (i.e., if a station is systematically picked late) will be absorbed into the source-specific station terms used in our initial locations. A possible future improvement to our method would be to stack the waveforms within each similar event cluster to create composite records with better signal to noise, which could then either be repicked or cross



**Figure 7.** Histograms of depth distributions for six different southern California seismicity catalogs.

correlated with composite records from other similar event clusters (provided they are sufficiently similar), as described by *Rowe et al.* [2004]. Potentially, this could improve the absolute locations of the clusters, but the relative locations would not be affected unless the centroid shifted by a considerable distance.

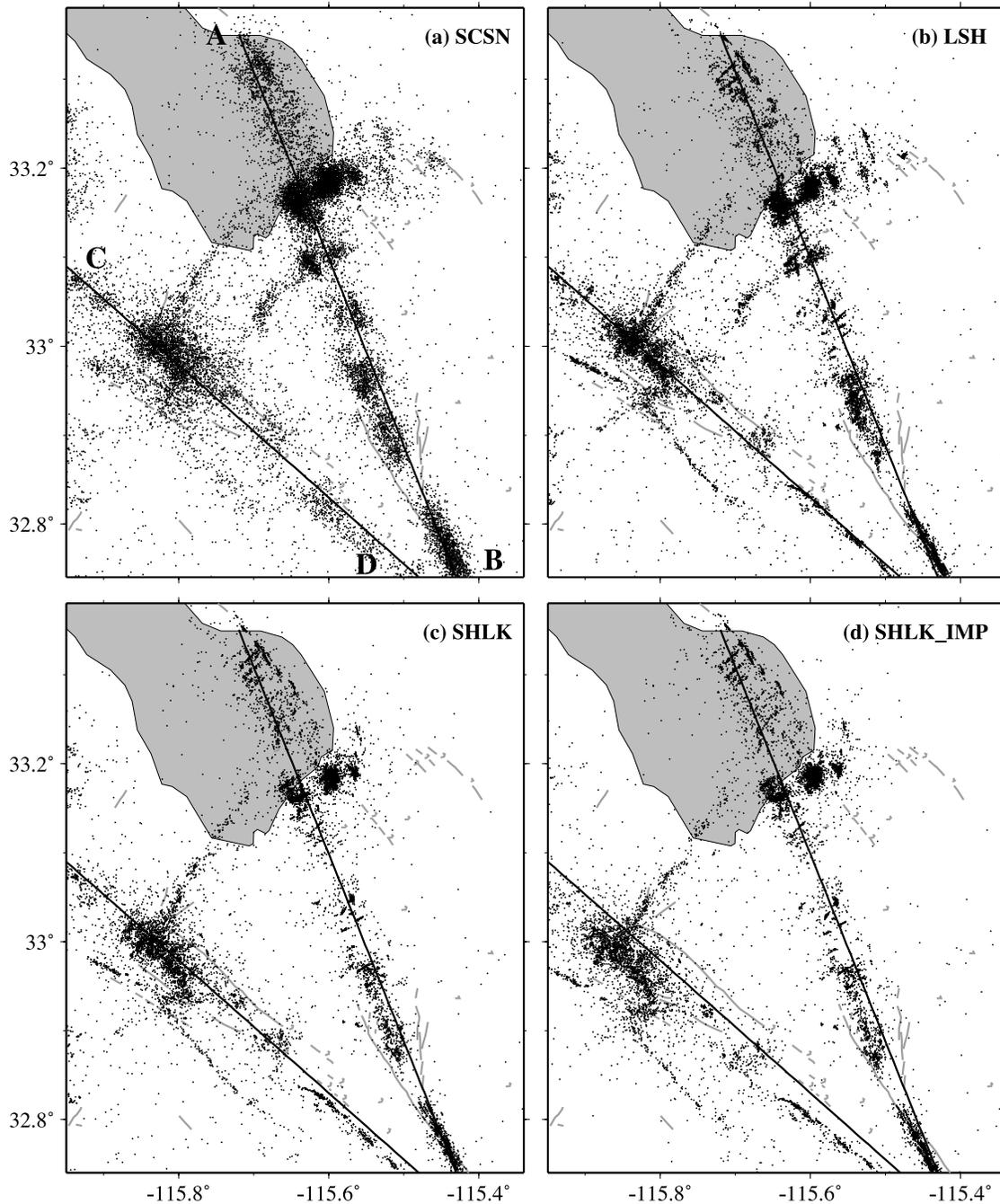
[19] For computational reasons, to compute the theoretical times for these differential locations, we use a 1-D velocity model (Figure 5), which is the layer-averaged model from the new 3-D model of *Lin et al.* [2007]. However, because this 1-D model is used only for differential locations based on differential times, most of the biasing effects of 3-D velocity structure are removed and we expect that very similar results would be obtained using a 3-D model. To test the sensitivity of our differential locations to changes in the velocity model, we computed results for some clusters using a different 1-D velocity model (that of *Shearer et al.* [2005]) and found only very small differences to the results presented here. The absolute locations of each cluster, of course, remain fixed at the cluster centroids as determined by ray tracing through our 3-D velocity model.

[20]  $M \geq 4$  earthquakes generally do not cross correlate well with smaller events because of their more complicated waveforms and frequent clipping of their records. However, we found that often our automated processing method would assign these large events to a similar event cluster and relocate them based on a small number of spurious

cross correlations with other events. We do not believe these locations are reliable and therefore we replaced the locations for the 896 local  $M \geq 4$  events in our final catalog with their 3-D locations obtained using the shrinking box SSST method. It is possible that more careful culling of the waveform data would allow these events to be relocated to greater accuracy but we do not attempt this here.

## 6. Relocation Results

[21] Figure 6 maps the final relocated seismicity for about 430,000 events from 1981 to 2005. Black dots show similar event clusters relocated using cross-correlation data. The distribution of similar event clusters is similar to those of *Shearer et al.* [2005]. About 25% of events do not correlate within clusters of at least 5 events and are plotted in color by year at their 3-D locations (24%) or 1-D locations (1%) if the 3-D locations are not available. For comparison, about 60% of the events in the SHLK catalog are in similar event clusters. According to *Schaff and Waldhauser* [2005], approximately 95% of the northern California seismicity includes events that have cross-correlation coefficients greater than 0.7 with at least one other event recorded at four or more stations. While the fraction of similar events depends to some extent on the details of the waveform similarity criteria required to define similar event pairs, it does appear that southern California seismicity is less likely



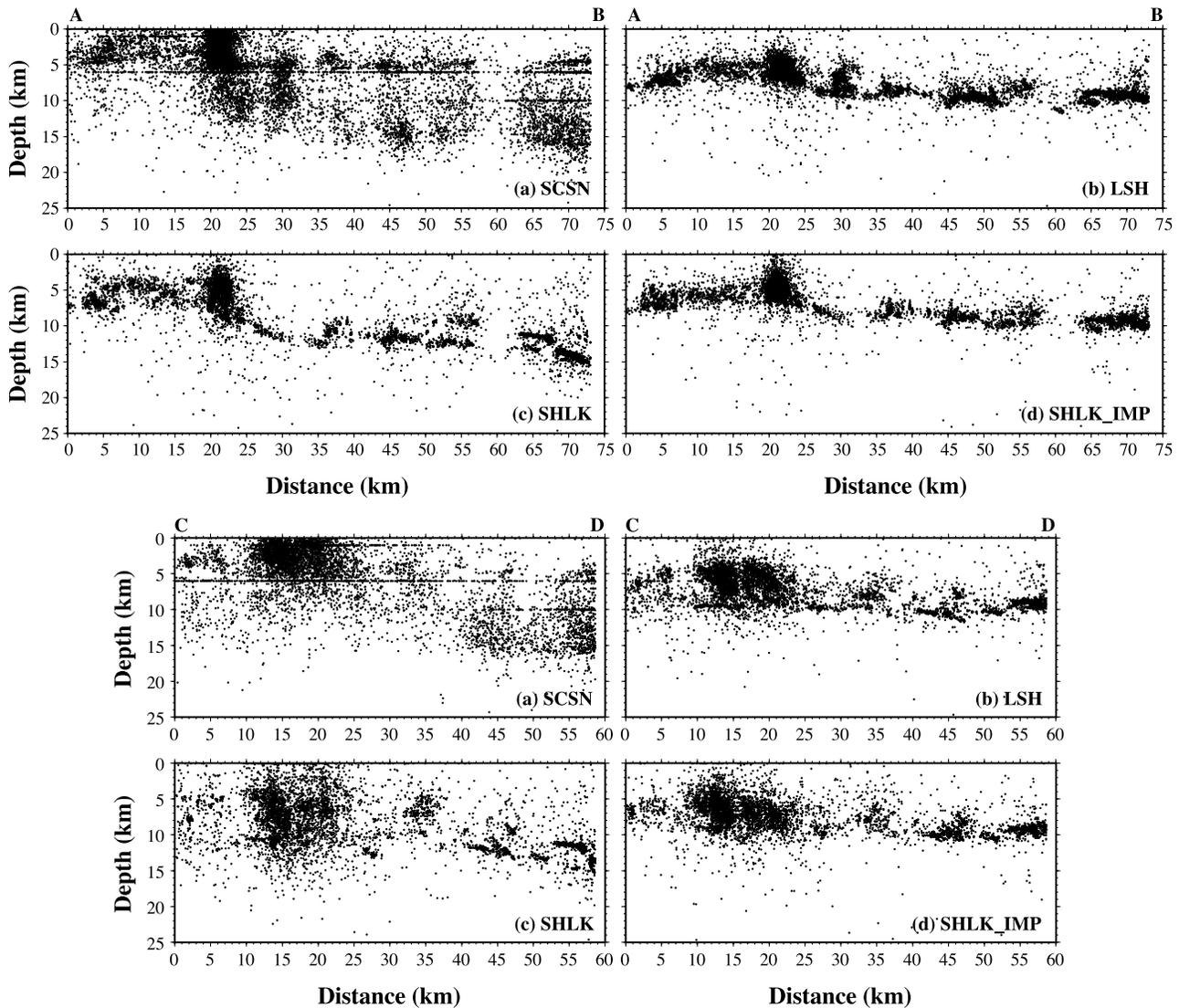
**Figure 8.** Epicenter distributions for the Imperial Valley region from (a) the SCSN catalog; (b) the new LSH catalog in this study; (c) the SHLK catalog using the 1-D model for southern California [see *Shearer et al.*, 2005, Figure 1a]; (d) the SHLK\_IMP catalog using the 1-D model derived from refraction seismic experiments [see *Shearer et al.*, 2005, Figure 1b]. The SCSN and LSH plots include events from 1981 to 2005; the SHLK plots include events from 1984 to 2002. The black straight lines are the profiles for the cross sections shown in Figure 9.

to occur in similar event clusters than northern California seismicity.

## 7. Comparisons Between Location Catalogs

[22] In this section, we compare locations of local events from six recent catalogs for southern California seismicity, including (1) the standard catalog of 408,105 local events

from 1981 to 2005 located by the SCSN using a layered 1-D velocity model for southern California; (2) the *Richards-Dinger and Shearer* [2000] catalog (referred as the RDS catalog) including 288,912 events from 1981 to 1998 relocated using a 1-D gradient velocity model and the SSST method; (3) the *Hauksson* [2000] catalog (referred as the HAUKSSON\_3D catalog) including 342,112 events from 1981 to 2000 relocated using a 3-D velocity model for



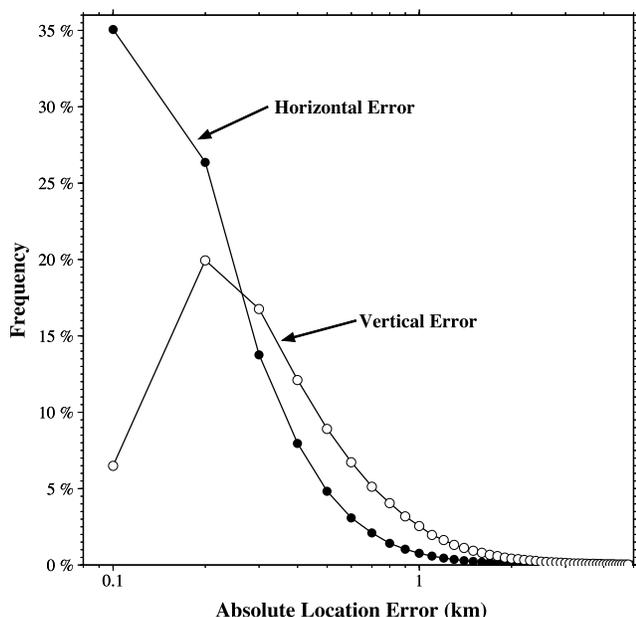
**Figure 9.** Cross sections of the seismicity within 10 km of the profiles shown in Figure 8 in the four location catalogs. The SCSN assigns a depth of 6 km to hypocenters with poorly constrained focal depths.

southern California; (4) the *Hauksson and Shearer* [2005] catalog (referred as the HAUKSSON\_DD catalog) including 327,430 events from 1984 to 2002 relocated using the double-difference (DD) location method [Waldhauser and Ellsworth, 2000; Waldhauser, 2001] and waveform cross correlation based on the 3-D initial locations [Hauksson, 2000]; (5) the SHLK catalog by *Shearer et al.* [2005] including 316,020 events from 1984 to 2002 relocated using the 1-D SSST method, similar event cluster analysis and waveform cross-correlation data; and (6) the new locations presented in this study (referred as the LSH catalog) including 399,521 events from 1981 to 2005 using cluster analysis, waveform cross-correlation data and a new robust differential time location method based on 3-D starting locations.

### 7.1. Depth Distribution

[23] First, we compare histograms of depth distributions from the six catalogs in Figure 7. In the SCSN catalog, there

is a big peak at 6 km depth, which may be due to the velocity discontinuity at 6 km depth in the layered velocity model. The RDS catalog was obtained based on a 1-D gradient velocity model and a SSST relative location method. Because of the relatively fast near-surface velocity in this model, this catalog has more events located between 0 and 3 km depth. For the HAUKSSON\_3D catalog, because of the simultaneous inversion of earthquake locations and velocity perturbations, the distribution of depths is very smooth. On the basis of this catalog, the relative locations are refined by the double-difference location method and waveform cross-correlation data in the HAUKSSON\_DD catalog. There is a small peak in this catalog at 6 km depth, which might be due to the sharp change in the velocity slope at 6 km depth in the 1-D velocity model [see *Hauksson and Shearer*, 2005; *Shearer et al.*, 2005]. There is also a small drop at about 9 km, which might be caused by the inconsistency between the 3-D and 1-D velocity models used for the 3-D and DD



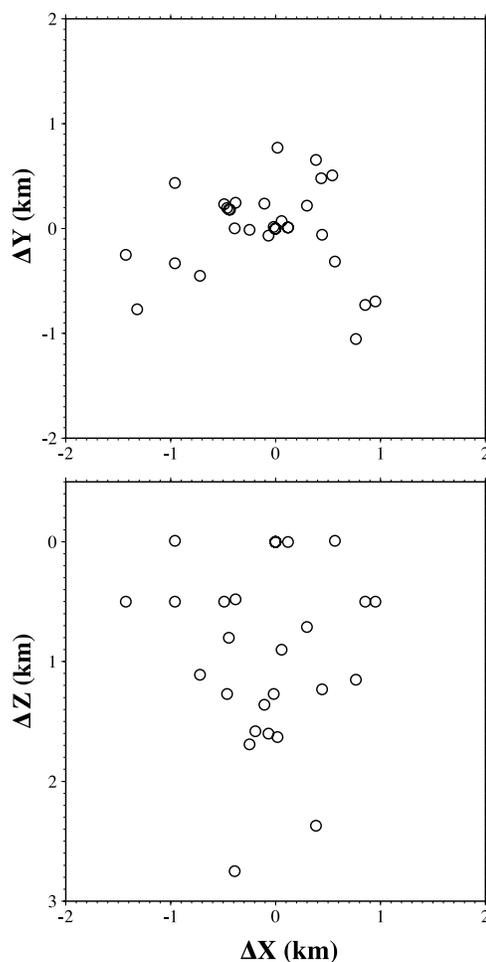
**Figure 10.** Histograms of absolute hypocenter errors from the SIMULPS results for the 430,000 relocated events. The median is 0.2 km for the horizontal errors and 0.4 km for the vertical errors.

locations. The distribution of the SHLK catalog is similar to the RDS catalog for depths more than 10 km and has fewer shallower events. Our new LSH location catalog has depths similar to that of both HAUSSON\_DD and SHLK. At shallow depths (<6 km), it is similar to the HAUSSON\_DD, whereas deeper it is similar to SHLK. This is reasonable because our new locations are based on a 3-D velocity model (as in the HAUSSON\_DD), cluster analysis and waveform cross correlation (same as in the SHLK). However, the distribution in our new catalog is more uniform, especially for shallower depths.

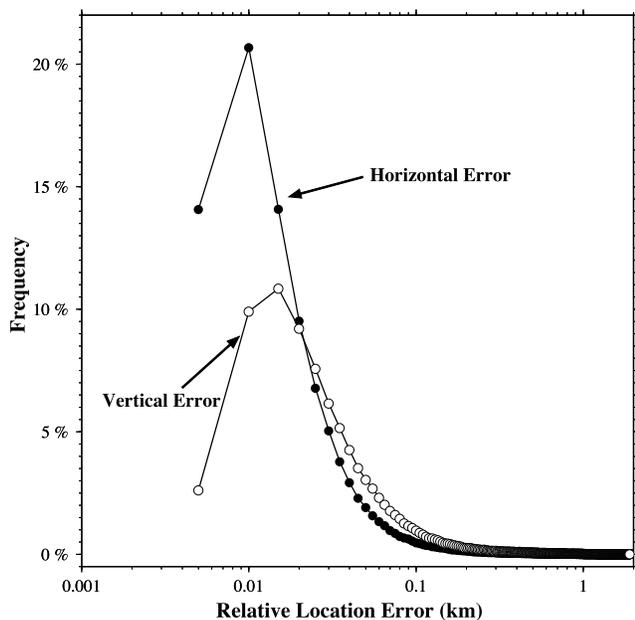
## 7.2. Imperial Valley Region

[24] In our previous catalog [Shearer *et al.*, 2005], we used a custom 1-D velocity model to obtain a more accurate set of hypocenters for the Imperial Valley, due to the substantially different velocity structure in this large sedimentary basin compared to the rest of southern California. In order to test the effectiveness of our 3-D model in accounting for this anomalous structure, we compare both the epicenter and depth distributions for this region from four catalogs. Figure 8 shows the map view of the seismicity for this region from the SCSN catalog (Figure 8a), the LSH catalog in this study (Figure 8b), the SHLK catalog using the 1-D model for southern California [see Shearer *et al.*, 2005, Figure 1a] (Figure 8c), and the SHLK\_IMP catalog using the 1-D model derived from refraction seismic experiments [see Shearer *et al.*, 2005, Figure 1b] (Figure 8d). All the events that fall within the box are plotted. More events are shown in Figures 8a and 8b because the LSH catalog extends from 1981 to 2005 while the SHLK catalog only extends from 1984 to 2002. The black straight lines are the profiles for the cross sections shown in Figure 9. In Figures 8b, 8c, and 8d, the relocated seismicity is sharper

than in the standard SCSN catalog. We observe differences in absolute locations for some clusters in this region between the LSH catalog and the SHLK and SHLK\_IMP catalogs that are due to the 3-D ray tracing; however, the relative locations are very similar. The absolute locations in the new LSH catalog are preferred because only ray tracing through a 3-D model can correctly account for the strong lateral velocity changes at the edges of the Imperial Valley. The biases from different 1-D velocity models are likely to be strongest in depth. Figure 9 presents the cross sections of the seismicity within 10 km of the profiles shown by the straight lines in Figure 8. From 27 to 32 km in the AB profile of the LSH catalog, there is a large blob of hypocenters shallower than those plotted in the SHLK or SHLK\_IMP catalogs. These events do not appear in SHLK because they are in the six additional years covered by LSH. The relocated seismicity is much sharper than the SCSN catalog. The new locations and those in the SHLK\_IMP catalog appear more stable than those in the SHLK catalog and more tightly clustered. At around both 55 km and 65 km in the AB profile there are two short horizontal stripes of



**Figure 11.** Location errors for the 36 shots used in the tomographic inversions and relocated in the new 3-D velocity model. There are 32 and 28 events in the  $\Delta x - \Delta y$  and  $\Delta x - \Delta z$  plots, respectively. Other events are not plotted because their errors are larger than the limits of the plots.



**Figure 12.** Histograms of relative location errors from bootstrap resampling for the 323,000 cross-correlation relocated events.

seismicity in the SHLK catalog that nearly collapse onto one another in both the SHLK\_IMP and LSH catalogs. The separation in SHLK is likely an artifact of an inappropriate velocity model. Note the similarity between the new LSH catalog and the SHLK\_IMP catalog, which suggests that the 3-D velocity model produces reasonably unbiased absolute locations. We therefore do not apply a separate velocity model for the Imperial Valley region in our new study.

## 8. Location Error Estimates

[25] Our new locations produce a dramatic sharpening of seismicity features compared to standard catalogs and a significant sharpening of some features compared to our previous SHLK catalog. Presumably this sharpening indicates decreased location errors because there is nothing intrinsic to our algorithms that should cause linear seismicity alignments. However, it is desirable to compute quantitative estimates of likely location errors for individual events. In order to do this, we estimate the absolute and relative location errors separately.

### 8.1. Absolute Location Errors

[26] The SIMULPS algorithm [Evans *et al.*, 1994] provides the hypocenter error ellipse. It computes the errors as the largest of the horizontal and vertical projections of the principal standard errors for each single event. Because we keep the absolute location of each similar event cluster fixed during the waveform cross-correlation relocation, we use the hypocenter errors from the output of the 3-D relocation to provide estimates of the absolute location errors in our catalog. The absolute location error in horizontal and vertical is given for each single event. We plot histograms of the hypocenter errors for the 430,000 events in Figure 10. The median of the horizontal errors is 0.2 km, and 0.4 km

for the vertical errors. These should be considered minimum errors because they represent the formal statistical errors in the solution and do not fully account for the possibility of errors in the velocity model or other systematic biases.

[27] We also used the 3-D velocity model to independently relocate the hypocenters and origin times of the 36 shots with known locations and origin times used by Lin *et al.* [2007]. Figure 11 shows the location errors for these shots. Except for 3 of the 36 events, all events have epicenter errors less than 1.5 km, with most errors less than 1.0 km. The 3 exceptional events have epicenter errors of about 2.8 km. For the vertical location errors, 30 are less than 3.0 km, and with most are less than 2.0 km. The others are about 6.0 km. For the 36 shots used here, the number of *P* picks is 1349, but the number of *S* picks is only 19. Considering this aspect and also the rapid velocity variations in the near surface, the absolute location errors from these shots can be treated as the maximum likely errors in our location catalog.

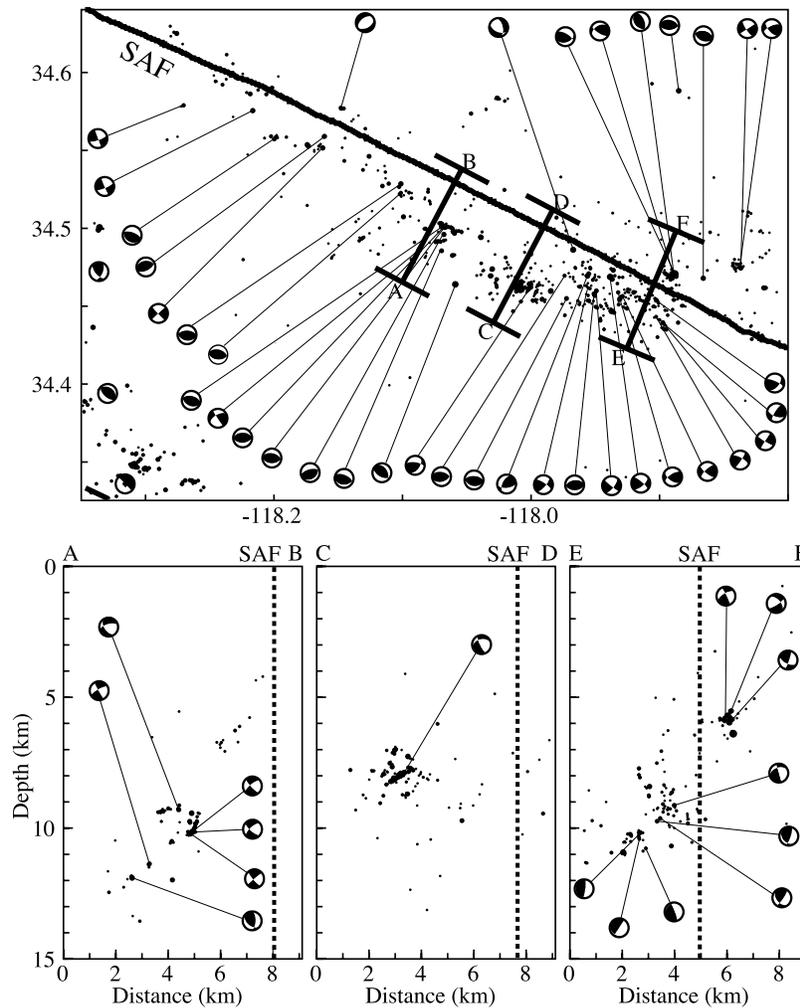
### 8.2. Relative Location Errors

[28] Within each similar event cluster, we apply a grid search algorithm using the robust least squares method to relocate all the events. Because we do not use the  $L_2$  norm, it is not possible to compute error ellipses based upon the  $\chi^2$  misfit criteria of the classical least squares method. As an alternative, we have applied a bootstrap approach [Efron and Gong, 1983; Efron and Tibshirani, 1991], in which the differential times for each event are randomly resampled (individual times may be sampled multiple times or not sampled at all). This process is repeated for 20 subsamples for each event and we relocate each event using the resampled differential times. We estimate the standard deviations of these 20 subsamples as the standard errors of the relative locations for each event. However, it should be noted that these formal statistical uncertainties can be quite small when the number of data points is large. Again, we plot the histograms of the relative errors in both horizontal and vertical locations from 323,000 cross-correlation relocated events in Figure 12. The median is 16 m for the relative horizontal location error and 34 m for the vertical location error. More accurate individual error estimates could be obtained by performing more than 20 bootstrap resamplings (limited computing time prevented us doing this), but the overall statistical properties of the estimated errors over many events should be reliably obtained even with a limited number of resamplings. It is possible that a faster method of estimating relative location errors could be obtained by more direct methods, such as studying the size and shape of the individual event misfit functions, but we do not attempt this here.

[29] In general, our estimated absolute location errors are similar to the HAUKSSON\_3D catalog and the relative location errors are comparable to the SHLK catalog.

## 9. Discussion

[30] Our new catalog is available through the Southern California Earthquake Data Center and we anticipate that it will be useful to a variety of researchers studying seismicity and tectonics in southern California. In particular our improved locations should help to resolve details of fine-

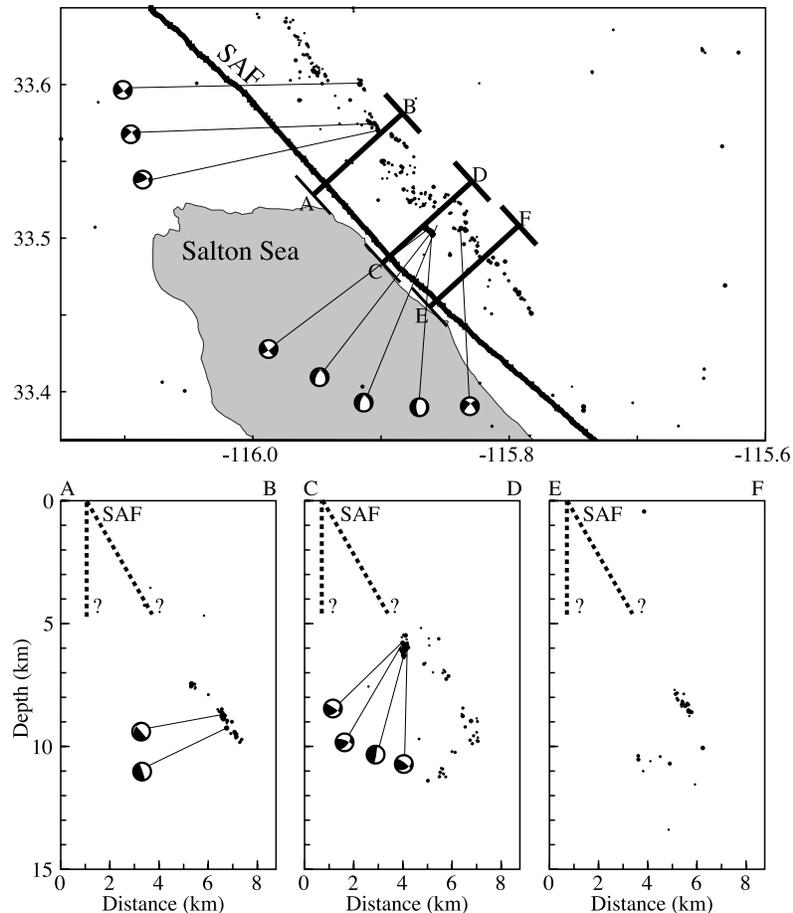


**Figure 13.** Seismicity near the San Andreas Fault (SAF) north of the San Gabriel Mountains, shown in map view and the labeled cross sections. Focal mechanisms are from the Hardebeck 2005 catalog (see text). Lower hemisphere focal spheres are plotted in the map view; far-side focal spheres are plotted in the cross sections. The dashed lines show the vertical projection of the known surface trace of the SAF; the true position of the fault at depth is uncertain.

scale fault structure in regions of active seismicity. One important question is the exact relationship between small earthquakes and major faults [e.g., *Hauksson et al.*, 2006]. Is the seismicity near these faults actually on the fault surface or on nearby subsidiary faults? A comprehensive study of this question is beyond the scope of this paper. However, we will highlight results from our catalog for two portions of the San Andreas Fault (SAF).

[31] In general there is sparse seismicity on the SAF in southern California, particularly when compared to the linear seismicity features that characterize the SAF in central California. However, there are two regions where groups of earthquakes are close to the SAF and aligned roughly parallel to the fault. One of these is between Palmdale and Wrightwood along the northern side of the San Gabriel Mountains. Figure 13 shows a map view and cross sections of earthquake locations in the western part of this region, where the seismicity is densest. Focal mechanisms are plotted from the quality 1 and 2 solutions of J. Hardebeck (2005, <http://www.data.scec.org/research/>

[social\\_focal\\_JLH.html](http://www.data.scec.org/research/social_focal_JLH.html); see also *Hardebeck and Shearer* [2003]). The bulk of the seismicity is located several kilometers south of the surface expression of the SAF, a much larger difference than our estimated location errors. The orientation of the SAF at depth is not known but is often assumed to be vertical (Community Fault Model, updated January 2004, by J. Shaw et al., <http://structure.harvard.edu/cfm>, also see *Plesch et al.* [2002]). The mechanisms are mainly reverse, oblique and right-lateral strike slip, with relatively few mechanisms matching exactly the expected motion along the SAF. In cross section the seismicity often appears to roughly align on southwest dipping planes. The focal mechanisms for cross sections AB and CD are consistent with reverse faulting on these planes. The seismicity is more complex in cross section EF where the events are closer to the SAF surface trace. Earthquakes located south of the fault deepen to the southwest but have mainly vertical, strike-slip mechanisms. It is possible that these events are located on strike-slip faults parallel to the SAF or even on the SAF itself if the



**Figure 14.** Seismicity near the San Andreas Fault (SAF) north of the Salton Sea, shown in map view and the labeled cross sections. Focal mechanisms are from the Hardebeck 2005 catalog (see text). The labeled line shows the known surface trace of the SAF; the true position of the fault at depth is uncertain.

fault dips slightly to the southwest at depth. However, there is also a concentration of seismicity at 6 km depth just north of the SAF, which contains reverse faulting mechanisms of varying orientation.

[32] The second region is northeast of the Salton Sea along the southernmost section of the SAF and is plotted in Figure 14. The seismicity is located 3 to 5 km to the northeast of the surface trace of the SAF, a separation that greatly exceeds any likely errors in our locations. Most of the earthquakes are between 6 and 11 km deep and roughly align on northeast dipping planes whose surface projection is close to the SAF. High-quality focal mechanisms are sparse in this region but the majority suggest right-lateral strike-slip motion, oriented parallel to the SAF. An exception is a concentration of seismicity in cross section CD, which contains 3 normal faulting mechanisms. However, the bulk of the seismicity could be occurring on the SAF if it dips about 60 degrees to the northeast. This geometry could also help explain geodetic data for this area, which indicate that the maximum shear strain is displaced about 7 km northeast of the surface trace of the SAF [Fialko, 2006].

[33] Our new earthquake locations are often sufficiently precise within similar event clusters that seismicity planes can be identified at relatively small scales (typical examples are from 0.5 to 2 km across). This resolution should permit

much more detailed mapping of fault geometries than has previously been possible using catalogs derived from more standard earthquake location methods, at least for those faults currently illuminated by seismicity. The orientations of these planes can also help resolve the ambiguity between the primary and auxiliary planes in focal mechanism solutions [Shearer *et al.*, 2003].

## 10. Conclusions

[34] We present high-precision earthquake locations for southern California from 1981 to 2005 computed using waveform cross correlation with a new robust least squares method. We examine 6 more years of data and many more cross-correlated event pairs in this study relative to our previous catalog. We use a new 3-D velocity model to improve absolute location accuracy and apply a new differential time relocation method that is very robust to outliers in the data. The location error estimates provide information on the location quality for individual events and the overall data set. These results build on our earlier relocation work and provide additional insight regarding the fine-scale seismicity structure in southern California. Our catalog is available through the Southern California Earthquake Data Center. Ultimately our goal is to implement these methods into routine network practice so that future events can be

located quickly to the same accuracy as the complete catalog.

[35] **Acknowledgments.** We thank Cliff Thurber for his SIMULPS tomography package and Paul Wessel and Walter Smith for developing and supporting the GMT mapping tools (<http://gmt.soest.hawaii.edu/>). We also thank Steve Cohen, Keith Richards-Dinger, and an anonymous reviewer for their detailed and constructive reviews. 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. This is SCEC contribution 1064.

## 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, Mod. Approaches Geophys.*, vol. 18, edited by C. Thurber and N. Rabinowicz, pp. 231–263, Kluwer Acad., Dordrecht, Netherlands.
- Astiz, L., and P. M. Shearer (2000), Earthquake locations in the inner continental borderland, offshore southern California, *Bull. Seismol. Soc. Am.*, *90*, 425–449.
- Astiz, L., P. M. Shearer, and D. C. Agnew (2000), Precise relocations and stress change calculations for the Upland earthquake sequence in southern California, *J. Geophys. Res.*, *105*, 2937–2953.
- 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.
- Eberhart-Phillips, D. (1990), Three-dimensional  $P$  and  $S$  velocity structure in the Coalinga region, California, *J. Geophys. Res.*, *95*, 15,343–15,363.
- Efron, B., and G. Gong (1983), A leisurely look at the bootstrap, the jack-knife and cross-validation, *Am. Stat.*, *37*, 36–48.
- Efron, B., and R. Tibshirani (1991), Statistical data analysis in the computer age, *Science*, *253*, 390–395.
- Evans, J. R., D. Eberhart-Phillips, and C. H. Thurber (1994), User's manual for SIMULPS12 for imaging  $V_p$  and  $V_p/V_s$ : A derivative of the "Thurber" tomographic inversion SIMUL3 for local earthquakes and explosions, *U. S. Geol. Surv. Open File Rep.*, 94-431.
- Felzer, K. R., and E. E. Brodsky (2006), Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature*, *441*, 735–738.
- Fialko, Y. (2006), Interseismic strain accumulation and earthquake potential on the southern San Andreas fault system, *Nature*, *441*, 968–971.
- Gillard, D., A. M. Rubin, and P. Okubo (1996), Highly concentrated seismicity caused by deformation of Kilauea's deep magma system, *Nature*, *384*, 343–346.
- 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.
- Hardebeck, J. L., and P. M. Shearer (2003), Using  $S/P$  amplitude ratios to constrain the focal mechanisms of small earthquakes, *Bull. Seismol. Soc. Am.*, *93*, 2434–2444.
- Hauksson, E. (2000), Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California, *J. Geophys. Res.*, *105*, 13,875–13,903.
- Hauksson, E., and P. Shearer (2005), Southern California hypocenter relocation with waveform cross-correlation, part 1: Results using the double-difference method, *Bull. Seismol. Soc. Am.*, *95*, 896–903.
- Hauksson, E., J. Woessner, and P. Shearer (2006), Associating seismicity to late Quaternary faults in southern California, *Eos Trans. AGU*, *87*(52), Fall Meet. Suppl., Abstract S11B-04.
- Huber, P. J. (1973), Robust regression: asymptotics, conjectures and monte carlo, *Ann. Stat.*, *1*, 799–821.
- Kagan, Y. Y. (2006), Earthquake spatial distribution: The correlation dimension, *Geophys. J. Int.*, *168*, 1175–1194, doi:10.1111/j.1365-246X.2006.03251.x.
- Lin, G., and P. Shearer (2005), Tests of relative earthquake location techniques using synthetic data, *J. Geophys. Res.*, *110*, B04304, doi:10.1029/2004JB003380.
- Lin, G., and P. M. Shearer (2007), Estimating local  $V_p/V_s$  ratios within similar earthquake clusters, *Bull. Seismol. Soc. Am.*, *97*, 379–388.
- Lin, G., P. M. Shearer, E. Hauksson, and C. H. Thurber (2007), A three-dimensional crustal seismic velocity model for southern California from a composite event method, *J. Geophys. Res.*, *112*, B11306, doi:10.1029/2007JB004977.
- Liu, J., K. Sieh, and E. Hauksson (2003), A structural interpretation of the aftershock "cloud" of the 1992  $M_w$  7.3 Landers earthquake, *Bull. Seismol. Soc. Am.*, *93*, 1333–1344.
- Lohman, R. B., and J. J. McGuire (2007), Earthquake swarms driven by aseismic creep in the Salton Trough, California, *J. Geophys. Res.*, *112*, B04405, doi:10.1029/2006JB004596.
- Moriya, H., H. Niitsuma, and R. Baria (2003), Multiplet-clustering analysis reveals structural details within the seismic cloud at the Soutz Geothermal Field, France, *Bull. Seismol. Soc. Am.*, *93*, 1606–1620.
- 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.
- Nakamura, Y. (1978),  $A_1$  moonquakes: Source distribution and mechanism, *Proc. Lunar Planet. Sci. Conf.*, *9*, 3589–3607.
- Nicholson, T., M. Sambridge, and O. Gudmundsson (2004), Three-dimensional empirical traveltimes: Construction and applications, *Geophys. J. Int.*, *156*, 307–328.
- Plesch, A., et al. (2002), SCEC 3D Community Fault Model for southern California, *Eos Trans. AGU*, *83*(47), Fall Meet. Suppl., Abstract S21A-0966.
- Powers, P., and T. Jordan (2005), Seismicity rate vs. distance from strike-slip faults in southern California, *Eos Trans. AGU*, *86*(52), Fall Meet. Suppl., Abstract S13B-0199.
- Richards-Dinger, K. B., and P. M. Shearer (2000), Earthquake locations in southern California obtained using source-specific station terms, *J. Geophys. Res.*, *105*, 10,939–10,960.
- Rowe, C., C. Thurber, and R. White (2004), Dome growth behavior at Soufriere Hills Volcano, Montserrat, revealed by relocation of volcanic event swarms, 1995–1996, *J. Volcanol. Geotherm. Res.*, *134*, 199–221.
- 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. Seismol. 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., and F. Waldhauser (2005), Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seismol. Soc. Am.*, *95*, 2446–2461.
- 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.
- Shearer, P. M. (1998), Evidence from a cluster of small earthquakes for a fault at 18 km depth beneath Oak Ridge, southern California, *Bull. Seismol. Soc. Am.*, *88*, 1327–1336.
- Shearer, P. M. (2002), Parallel fault strands at 9-km depth resolved on the Imperial Fault, Southern California, *Geophys. Res. Lett.*, *29*(14), 1674, doi:10.1029/2002GL015302.
- Shearer, P. M., J. L. Hardebeck, L. Astiz, and K. B. Richards-Dinger (2003), Analysis of similar event clusters in aftershocks of the 1994 Northridge, California, earthquake, *J. Geophys. Res.*, *108*(B1), 2035, doi:10.1029/2001JB000685.
- Shearer, P. M., E. Hauksson, and G. Lin (2005), Southern California hypocenter relocation with waveform cross-correlation, part 2: Results using source-specific station terms and cluster analysis, *Bull. Seismol. Soc. Am.*, *95*, 904–915.
- Thurber, C., and D. Eberhart-Phillips (1999), Local earthquake tomography with flexible gridding, *Comput. Geosci.*, *25*, 809–818.
- Thurber, C. H. (1983), Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, *J. Geophys. Res.*, *88*, 8226–8236.
- Thurber, C. H. (1992), Hypocenter-velocity structure coupling in local earthquake tomography, *Phys. Earth Planet. Inter.*, *75*, 55–62.
- Thurber, C. H. (1993), Local earthquake tomography: velocities and  $V_p/V_s$ -theory, in *Seismic Tomography: Theory and Practice*, edited by H. M. Iyer and K. Hirahara, pp. 563–583, Chapman and Hall, London.
- Waldhauser, F. (2001), hypoDD: A program to compute double-difference hypocenter locations (hypoDD version 1.0, 03/2001), *U. S. Geol. Surv. Open File Rep.*, 01-113.
- Waldhauser, F., and W. L. Ellsworth (2000), A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California, *Bull. Seismol. Soc. Am.*, *90*, 1353–1368.
- Waldhauser, F., W. L. Ellsworth, and A. Cole (1999), Slip-parallel seismic lineations on the northern Hayward Fault, California, *Geophys. Res. Lett.*, *26*, 3525–3528.
- Zhang, H., and C. H. Thurber (2003), Double-difference tomography: The method and its application to the Hayward Fault, California, *Bull. Seismol. Soc. Am.*, *93*, 1875–1889.

E. Hauksson, Seismological Laboratory, MS 252-21, California Institute of Technology, Pasadena, CA 91125, USA. (hauksson@gps.caltech.edu)

G. Lin, Department of Geology and Geophysics, University of Wisconsin-Madison, 111 Weeks Hall, 1215 W. Dayton Street, Madison, WI 53706, USA. (glin@geology.wisc.edu)

P. M. Shearer, Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA. (pshearer@ucsd.edu)