JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, B08311, doi:10.1029/2007JB005479, 2008

Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods

Felix Waldhauser[1] and David P. Schaff[1]

Received 1 November 2007; revised 7 March 2008; accepted 9 April 2008; published 15 August 2008.

[1] We simultaneously reanalyzed two decades (1984–2003) of the digital seismic archive of Northern California using waveform cross-correlation (CC) and double-difference (DD) methods to improve the resolution in hypocenter locations in the existing earthquake catalog generated at the Northern California Seismic Network (NCSN) by up to three orders of magnitude. We used a combination of ~3 billion CC differential times measured from all correlated pairs of events that are separated by less than 5 km and ~7 million *P* wave arrival-time picks listed in the NCSN bulletin. Data were inverted for precise relative locations of 311,273 events using the DD method. The relocated catalog is able to image the fine-scale structure of seismicity associated with active faults and revealed characteristic spatiotemporal structures such as streaks and repeating earthquakes. We found that 90% of the earthquakes have correlated *P* wave and *S* wave trains at common stations and that 12% are colocated repeating events. An analysis of the repeating events indicates that uncertainties at the 95% confidence level in the existing network locations are on average 0.7 km laterally and 2 km vertically. Correlation characteristics and relative location improvement are remarkably similar across most of Northern California, implying the general applicability of these techniques to image high-resolution seismicity caused by a variety of plate tectonic and anthropogenic processes. We show that consistent long-term seismic monitoring and data archiving practices are key to increase resolution in existing hypocenter catalogs and to estimate the precise location of future events on a routine basis.

Citation: Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, *113*, B08311, doi:10.1029/2007JB005479.

1. Introduction

[2] The Northern California Seismic System (NCSS; Figure 1a), which assimilates data from 13 seismic networks, records an average of ~20,000 earthquakes on 1200 channels each year with ~1 million seismograms being added to the digital archive every year since 1984. The majority of earthquakes occur in diverse and complex tectonic settings such as the San Andreas Fault system (SAF), representing the boundary between the Pacific and North American plate, the volcanic region of Long Valley Caldera (LVC), and the Mendocino Triple Junction (MTJ) at the intersection of the Gorda, North American, and Pacific plate (Figure 1a). In addition, large numbers of anthropogenic earthquakes are induced at the Geysers Geothermal Field (GGF) by geothermal production activites.

[3] Earthquakes recorded by the NCSS are routinely located at the Northern California Seismic Network (NCSN) on an event-by-event basis by a linearized inversion of the seismic phase arrival times (mostly Pg) picked from the seismograms [*Geiger*, 1910; *Klein*, 2002]. These parametric data is archived in what we refer to in the following as the NCSN catalog, even though some of the short-period phase data comes from other networks (i.e., NN, WR, CI, PG & UW). Inaccuracies in the phase picks and errors in the model used to predict the data cause hypocenter location uncertainties in the range of several hundred meters to a few kilometers, with depth more poorly constrained than the epicenter. These errors are many times larger than the spatial dimension of the earthquakes themselves (~10 m to 1000 m for M1–4 earthquakes), and hamper the study of a wide range of scientific problems concerning the physics of earthquakes, the structure and composition of the Earth's interior, and the seismic hazard of active faults.

[4] Here, we take advantage of the dense distribution of recorded events that accumulated across most of Northern California over the last few decades, the associated comprehensive and consistently archived digital seismograms and parametric data from the Northern California Earthquake Data Center (NCEDC), and the growth in storage and computing capacity over the last several years. Earthquakes that are close together in space, and have similar rupture mechanisms, produce similar waveforms at common stations [*Poupinet et al.*, 1984]. Cross-correlation methods can then measure differential phase arrival times between two such correlated earthquakes and a common station with subsample accuracy [e.g., *Poupinet et al.*, 1984; *Deichmann*

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA.

0148-0227/08/2007JB005479$09.00

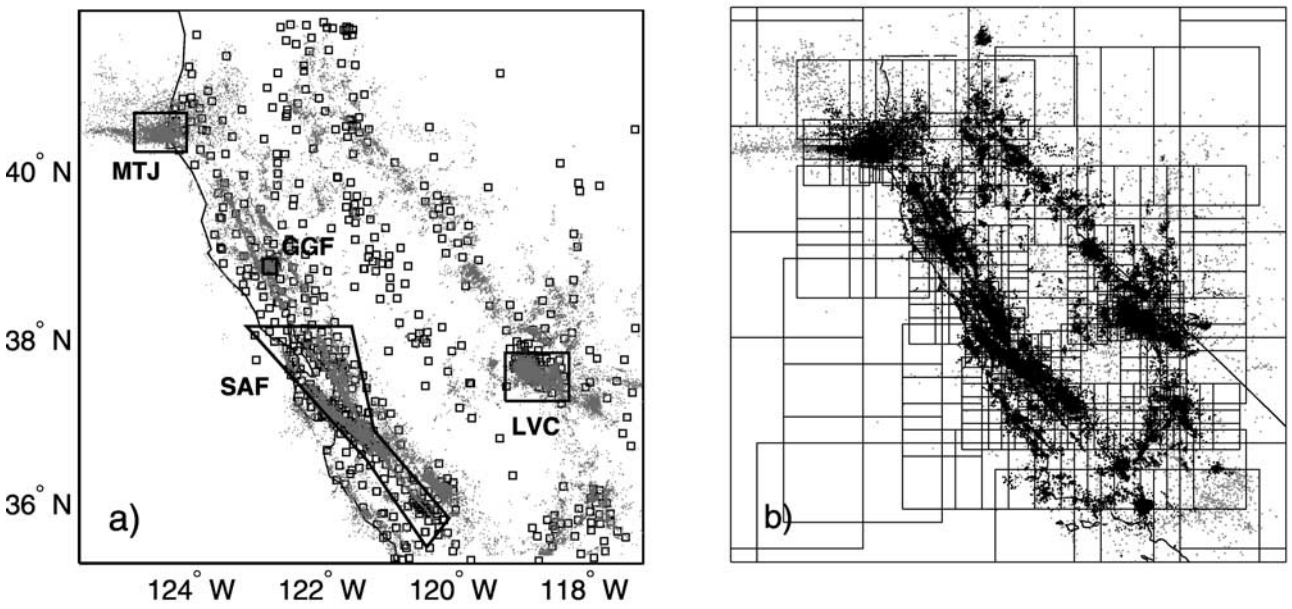

Figure 1. (a) Relocated earthquakes (n = 311,273; gray dots) in Northern California recorded by six or more stations (squares) of the Northern California Seismic System (NCSS) between January1984 and May 2003. Labeled polygons show the four focus regions discussed in this study: SAF, selected area of the San Andreas Fault system, including the San Andreas, Hayward, and Calaveras faults; MTJ, Mendocino Triple Junction; LVC, Long Valley Caldera; GGF, Geysers Geothermal Field. (b) Original earthquake locations (n = 408,084; gray dots) recorded and located by the NCSS. Black dots denote the 311,273 earthquakes (recorded by six or more stations) relocated in this study (plotted at their original NCSN location). Black boxes indicate areas used to parallelize the relocation procedure. Thin lines denote the California coastline and state border.

and Garcia-Fernandez, 1992; *Schaff et al.*, 2004]. Such differential arrival times can then be simultaneously inverted for the precise distance between events [e.g., *Got et al.*, 1994; *Waldhauser and Ellsworth*, 2000].

[5] Waveform-based multievent relocation methods have been used in numerous previous studies to minimize pick and model errors in an effort to increase the spatial resolution in routinely produced local earthquake locations [e.g., *Poupinet et al.*, 1984; *Deichmann and Garcia-Fernandez*, 1992; *Rubin et al.*, 1999; *Waldhauser et al.*, 1999; *Got and Okubo*, 2003; *Hauksson and Shearer*, 2005; *Shearer et al.*, 2005; *Richards et al.*, 2006]. Here we report on a large-scale, uniform and comprehensive application of efficient cross-correlation (CC) and double-difference (DD) methods to 20 years of the NCSS archive of waveform and parametric data. These techniques have been tested and applied in specialized studies in Northern California where they imaged detailed seismicity structures, shedding light on the mechanics of active faults [*Waldhauser et al.*, 1999, 2004; *Rubin et al.*, 1999; *Waldhauser and Ellsworth*, 2002; *Prejean et al.*, 2002; *Schaff et al.*, 2002]. The comprehensive nature of the study presented here permits us to quantify the general applicability of these techniques across diverse tectonic regions.

[6] In this paper we demonstrate orders of magnitude improvement in relative hypocenter locations over the currently available network catalog locations for most of Northern California. Absolute locations are also improved, although to a lesser degree, because of the improvement in relative locations, as long as there is no significant systematic bias in the centroid of a given cluster of network locations. This catalog and its future updates provide the fundamental data for many lines of research where the precision of hypocenter locations is critical, such as the study of fault structure and mechanics, earthquake statistics, the generation of earthquakes, and earthquake interaction. In this paper we focus on describing the new double-difference catalog, characterize and quantify the improvements over existing locations, investigate the underlying reasons that enable these improvements, and discuss the implications for seismic monitoring practices in general.

2. Seismic Data and Earthquake Relocation Procedure

[7] We use the entire seismic archive recorded between January 1984 and May 2003 by the NCSS and produced by the U.S Geological Survey and the University of California at Berkeley. This database, made available to us by the NCEDC (Doug Neuhauser, personal communication, 2001), includes 408,084 events, 15 million digital waveforms (800 Gb) recorded at ~500 short-period stations, and nearly 7 million P wave arrival-time picks.

2.1. Cross-correlation Measurements

[8] The cross-correlation differential times used in this study are based on a comprehensive analysis of the complete digital seismogram database for all pairs of events with hypocentral separations less than 5 km [*Schaff and Waldhauser*, 2005]. Event separations were computed after double-difference relocation of the NCSN catalog

Table 1. Statistics of the Cross-Correlation Based Double-Difference Catalog[a]

	ALL	SAF	LVC	MTJ	GGF
Number of relocated events (% with waveforms)	311,273 (66)	81,679 (82)	115,751 (53)	7,636 (85)	46,448 (53)
% correlated events	90	96	95	88	92
% repeating events	11.9	27.1	4.9	2	5.2
Median DD errors (*x*/*z*)	0.050/0.047	0.039/0.030	0.043/0.043	0.196/0.070	0.041/0.045
Mean DD errors (*x*/*z*)	0.450/0.290	0.308/0.068	0.199/0.103	1.025/0.227	0.090/0.071
Network errors (*x*/*z*)					
95% confidence level	0.715/2.069	0.749/1.675	0.513/1.662	1.552/2.361	0.280/1.818
Mean	0.172/0.257	0.158/0.243	0.138/0.208	0.340/0.325	0.096/0.190
Median	0.111/0.160	0.108/0.160	0.107/0.140	0.300/0.220	0.079/0.130
Maximum	7.983/6.958	3.670/6.958	2.048/3.218	2.242/2.600	0.887/1.947

[a]DD relative location errors are median and mean of the major axes of the horizontal and vertical projection of the 95% confidence ellipsoids calculated from 200 bootstrap samples for each event. Network relative location errors are estimated from an analysis of repeating events (Figure 8). All errors are in kilometers.

using phase picks alone. Vertical component seismograms with a sampling rate of 10 ms are available for 225,000 events, or 55% of the total of 408,084 events in the catalog between 1984–2003. Most of the events for which waveforms are not available at the NCEDC locate in the Long Valley and Geysers areas (see Table 1), and occurred in the early years of network operation when different trigger/waveform archiving mechanisms were in place. We relocate these events by just using the pick data.

[9] We applied a time-domain cross-correlation function [*Schaff et al.*, 2004] to 26 billion filtered (1.5–15 Hz) seismogram pairs, using both 1 s and 2 s windows around the *P* wave and *S* wave energy. Windows were initially aligned on the arrival-time picks when available, and on arrival times predicted using a simple layered 1-D model when no picks were made. All phase pairs with cross-correlation coefficients (Cf) greater than 0.6 were stored in a binary database, and a total of 1.7 billion *P* wave and

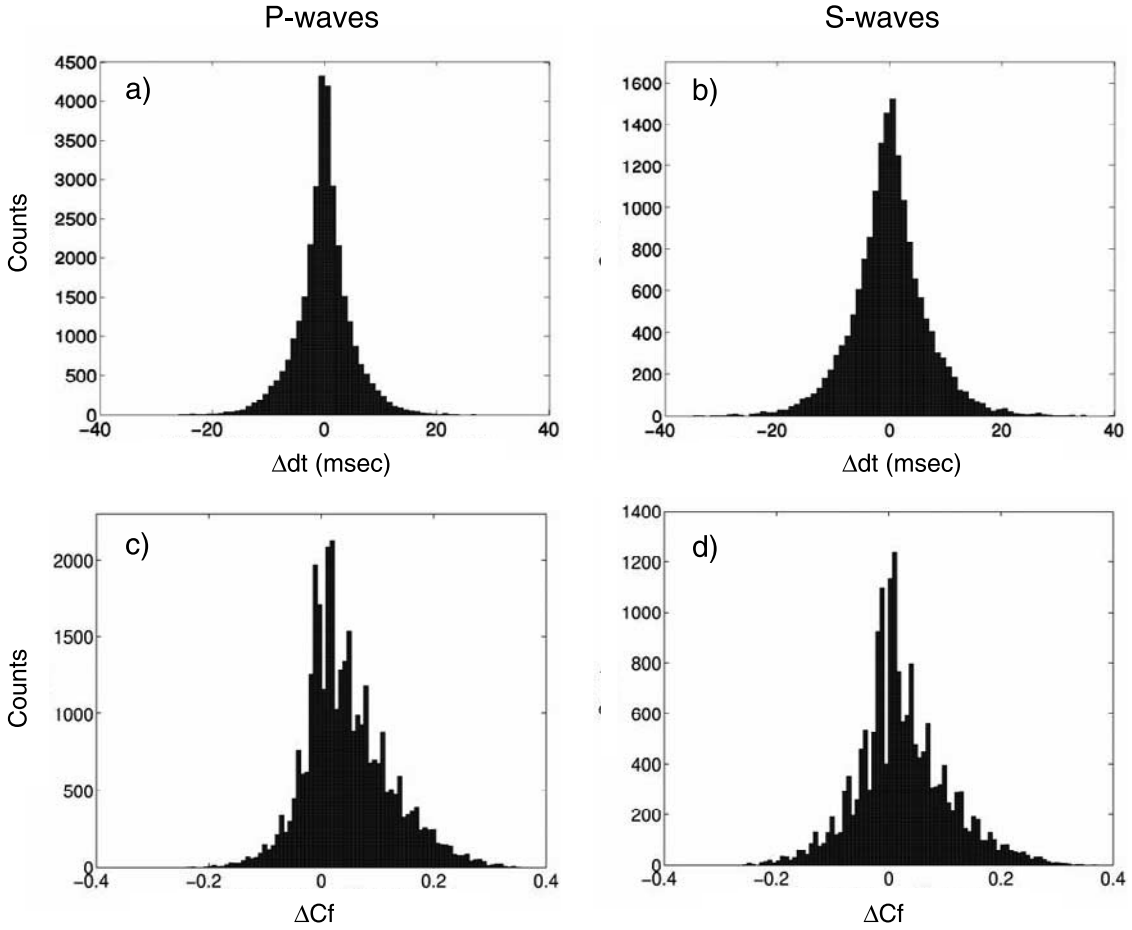

Figure 2. Histograms of differences between cross-correlation measurements obtained by *Schaff and Waldhauser* [2005] and P. Shearer (personal communication, 2004) for 3152 earthquakes near Mendocino, California. Differences are shown for (a) *P* wave and (b) *S* wave delay times and (c) *P* wave and (d) *S* wave correlation coefficients.

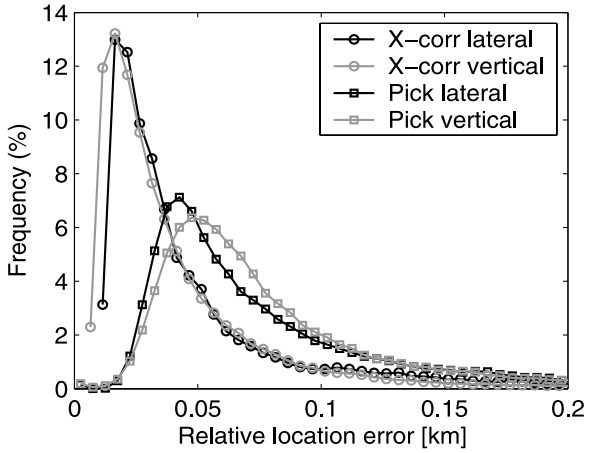

Figure 3. Histograms of lateral and vertical relative location errors, computed from the mayor axes of the horizontal and vertical projection of the 95% confidence ellipsoids obtained from a bootstrap analysis of the final double-difference vector based on 200 samples with replacement. Median values within bins of 0.005 km are shown for 123,035 event locations constrained by pick data only and for 181,414 locations constrained by at least four cross-correlation measurements.

1.2 billion *S* wave differential times were saved (see *Schaff and Waldhauser*, 2005, for details). Note that *S* waves have been rarely picked at the NCSN because picking S on vertical short-period stations is difficult and unreliable. This is about to change as the NCSN is migrating to digital stations with triaxial sensors where *S* waves can be picked from the horizontal component (D. H. Oppenheimer, personal communication, 2008).

[10] Inspection of delay time measurements with Cf < 0.7 indicates that a substantial amount of these measurements are outliers resulting from cycle skipping and the correlation of noise. Thus for the relocations presented in this study, we only use differential times with Cf ≥ 0.7. To further reduce outliers we inspect the consistency of the measurements obtained for the 1 and 2 s cross-correlation windows. Measurements with differences larger than one sample are removed. Correlated noise, for example, can be easily detected this way. We use delay times based on the 1 s window if the measurements pass the inspection. Finally, in order to assure robustness of the relocation process, we select all differential times from event pairs that have at least 4 P or S delay time measurements. This also helps to suppress outliers caused by time-dependent station delays, which may have high correlation coefficients but biased differential times [*Rubin*, 2002]. Such outliers are difficult to detect during the cross-correlation processing stage, but they typically show high residuals during relocation and can be detected and removed at this later stage.

[11] The outlier detection and data selection procedure results in a high-quality database of $\sim$200 million *P* wave and $\sim$100 million *S* wave differential time measurements that go into the relocation process. In areas with large numbers of highly correlated earthquakes a subset of the available cross-correlation differential times is selected in order to reduce the size of the system of double-difference equations. The selection criteria retains only a limited number of measurements per event pair (typically those with the highest correlation coefficients), and reduces the number of linked nearest neighbors without sacrificing optimal data connectivity between events.

[12] We compare our correlation data with independently obtained delay time measurements (Peter Shearer, personal communication, 2004) for 3152 earthquakes recorded at 32 NCSN stations near Mendocino, Northern California. Both our [*Schaff et al.*, 2004] and Shearer's [*Hauksson and Shearer*, 2005] method use a time-domain cross-correlation function, but employ different interpolation functions and cross-correlation parameters (e.g., window lengths, lags) that may cause differences in the delay time measurements. The differences between the two data sets are presented in Figure 2 for events common in both data sets. From a total of 17,684 differential times compared, 96% of the *P* wave data agree within 10 msec (the sampling rate), and 63% within 1 msec (Figure 2a). 92% of the *S* wave data agree within 10 msec, 59% within 1 msec (Figure 2b). Outliers are sparse, and present in both data sets. They are likely caused by glitches during the cross-correlation process such as cycle skipping or correlation of noise. A systematic shift toward higher correlation coefficients is apparent in our data because of our choice of a shorter window length (1 sec) compared to the one used by Shearer (2 sec for *P* waves, 3 sec for *S* waves) (Figures 2c and 2d).

2.2. Double-Difference Relocation

[13] We combine the cross-correlation differential times with $\sim$1 billion travel-time differences computed from $\sim$7 million NCSN P-phase arrival-time picks and relocate the NCSN catalog on a 64-processor Beowulf cluster with a modified version of the double-difference algorithm *hypoDD* [*Waldhauser*, 2001]. The double-difference method is an iterative least squares procedure that relates the residual between the observed and predicted phase travel-time difference for pairs of earthquakes observed at common stations to changes in the vector connecting their hypocenters through the partial derivatives of the travel times for each event with respect to the unknown [*Waldhauser and Ellsworth*, 2000]. This approach cancels common mode errors when the distribution of seismicity is sufficiently dense; i.e., where distances between neighboring events are small relative to station distances (typically a few kilometers or less). By linking hundreds or thousands of earthquakes together through a chain of near neighbors it is possible to obtain high-resolution relative hypocenter locations over a large area.

[14] From a total of 408,084 events in the NCSN catalog we choose 317,141 events for relocation that include event pairs with at least 6 phases observed at common stations to ensure robustness of the double-difference inversions (Figure 1b). We compute travel-time differences from NCSN picks between each event in the catalog to its 20 nearest neighbors within 10 km distance. Only 40 of the highest quality pick differential times per pair are selected. Both pick and cross-correlation differential times are combined in a dynamically weighted double-difference inversion to insure location precision of correlated events to the accuracy of the cross-correlation data, and of those that do not correlate (or have no archived waveforms) to the accuracy

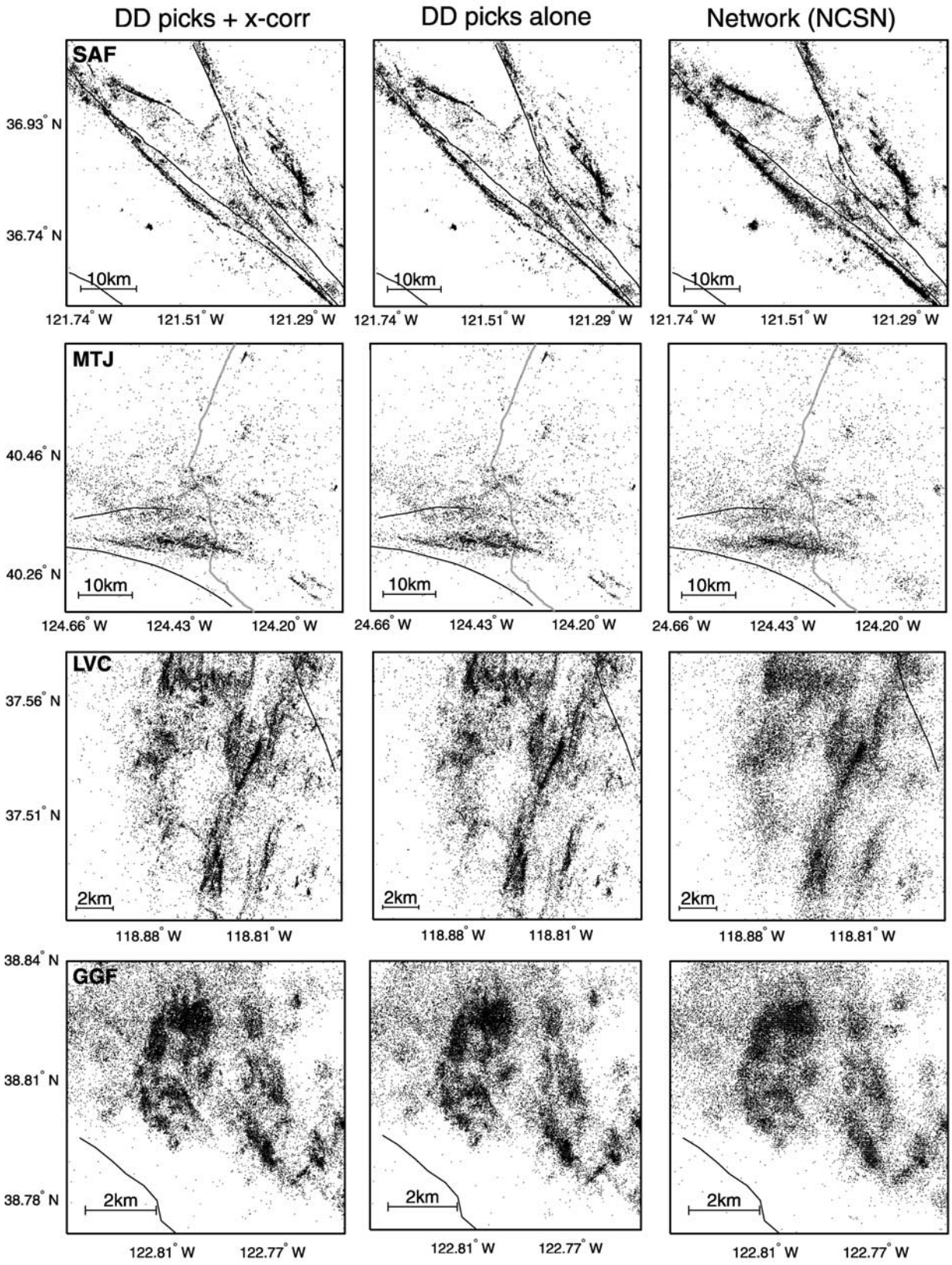

Figure 4. Comparison between cross-correlation (left panels) and pick-based (middle) double-difference locations and network locations (right panels) for representative subareas in each of the four focus regions (see Figure 1a for location of focus regions). The same events are displayed in each of the three panels associated with a focus region. Note the networks of discrete faults imaged in the relocated seismicity in the tectonic regions SAF, MTJ, and LVC, compared to sharpened ''clouds'' imaged in the region of induced seismicity at GGF. Lines indicate mapped surface fault traces.

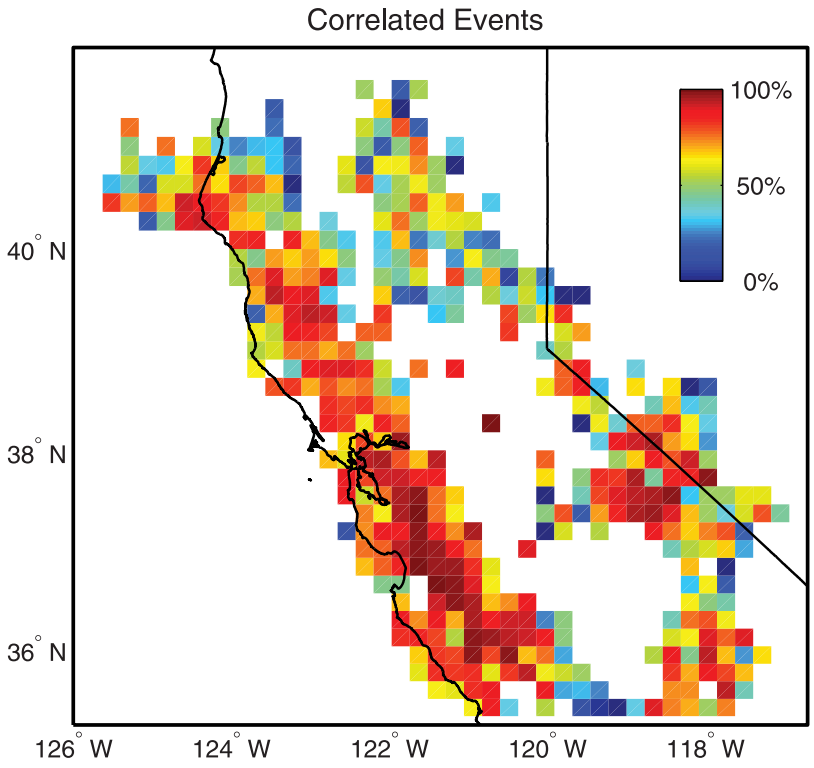

Figure 5. Percentage of correlated earthquakes across Northern California, displayed within cells of 20 × 20 km. Only cells with 10 or more events are shown. Solid lines denote coastline and state line.

of the pick data [*Waldhauser*, 2001]. The NCSN catalog locations are used as starting locations, and the locations and partial derivatives are being updated during each iteration.

[15] 1-D layered velocity models are used to locally predict travel times and partial derivatives. The models are chosen from a compilation of 39 local models used by the NCSN to locate the earthquakes one at a time on a routine basis [*Oppenheimer et al.*, 1993]. Most of these high-quality models were determined in separate studies by simultaneous inversion of seismic arrival times for changes in hypocenter locations and layer velocity, or established from local active source data. We resample the velocity-depth function of each model to generate models with 28 layers of constant velocity to avoid strong velocity jumps across interfaces. These models allow for a very efficient prediction of delay times and partial derivatives. In the discussion section below, we investigate the potential bias in the DD locations due to 3-D structures not represented by our 1-D models.

[16] We parallelize the relocation process, similar to *Hauksson and Shearer* [2005], by generating 513 rectangular boxes with square surface areas (Figure 1b). Each box includes events connected through a web of differential time links not exceeding 3 million. Size and location of the boxes are found by starting with one box that includes all events, repeatedly splitting the box up into smaller boxes until the number of links that connect the events within a given box falls below the maximum link threshold. Each box overlaps the area of its four neighboring boxes by 50%. We require a continuous chain of pair wise connected events with a link strength of 7 differential times. Differential times from stations within 150 km from an event pair's centroid are used.

[17] Modifications to the original *hypoDD* code (version 1.1; *Waldhauser*, 2001) include an automatic search for optimal double-difference parameters for each box, starting from initial parameters determined from experience on working with subsets of the NCSN catalog in selected areas in Northern California (see *Waldhauser*, 2001, for details on *hypoDD* parameters). Typically, each box undergoes a series of 20 iterations, during each of which the weighting of the delay time data is dynamically adjusted as a function of event separation and delay time residuals. The first ten iterations generally down-weight, by a factor of 0.01, the cross-correlation data to ensure robust first-order relative location improvement by minimizing model error bias via the pick data, and to avoid potential station bias associated with the cross-correlation measurements (see *Waldhauser*, 2001). The remaining ten iterations down-weight, by a factor of 0.01, the pick data in order to let the cross-correlation data resolve the fine details in relative event locations of correlated events. Proper damping of the LSQR [*Paige and Saunders*, 1982] solutions is automatically determined by investigating the condition number of the system of linear equations, and the rate of convergence.

[18] We relocate ~300,000 events in a few hours of time using an average of ten 1.2 GHz Athlon MP processors (excluding the time spent on establishing the differential time data sets). *hypoDD* output parameters (e.g., RMS, data outliers, convergence rate, etc.) summarizing performance and robustness of the DD solutions in each of the 513 relocation boxes are subsequently screened, and boxes with suspicious output values (less than 1%, mostly because of

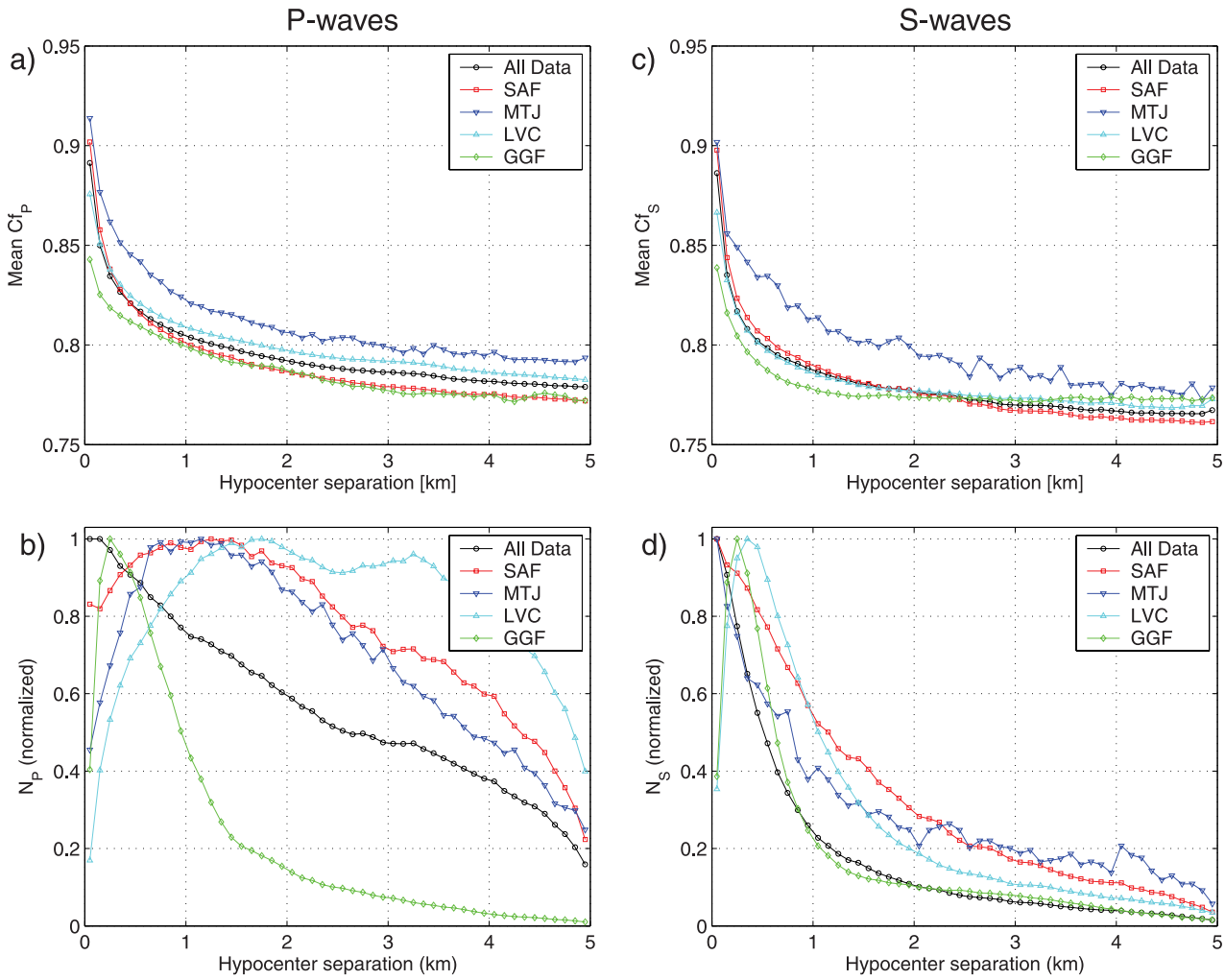

Figure 6. Distribution of (a, c) *P* wave and *S* wave cross-correlation coefficients and (b, d) normalized number of correlations shown as means within bins of 0.1-km distance between correlated events. *P* wave data in Figures 6a and 6b are based on 152 million *P* wave cross-correlation coefficients, with Cf $\geq$ 0.7 measured from 23 million pairs of correlated earthquakes; *S* wave data in Figures 6c and 6d are from $\sim$44 million *S* wave cross-correlation coefficients (Cf $\geq$ 0.7) derived from $\sim$7 million pairs of correlated events. Statistics are shown for all correlated events in Northern California (open circles) and for individual tectonic regions SAF, MTJ, LVC, and GGF. See Figure 1a for abbreviations and geographic locations.

numerical instabilities during inversion) inspected and reprocessed manually. The final double-difference solutions from each box are combined into a single catalog by forming a weighted location average of events that are included in more than one box. The weight is a linear function of an event's distance from the centroid of the cluster it belongs to.

3. Relocation Results

[19] The double-difference catalog includes 311,273 events between January 1984 and May 2003, or 98% of all events recorded at 6 or more stations in that time period (Figure 1a). Events are "lost" during the relocation process mostly because of insufficient data links after the weighting function removes outliers. The root mean square (RMS) of the weighted pick differential time residuals for the relocated events is 0.017 s, compared to 0.124 s before relocation. The weighted RMS of the cross-correlation data is 0.004 s after relocation. Relative location errors are estimated for each event by bootstrapping, with replacement, the final unweighted double-difference residual vector [*Waldhauser and Ellsworth*, 2000]. Error ellipsoids are obtained at the 95% confidence level for 200 bootstrap samples. The distribution of the major axes of the horizontal and vertical projections of these ellipsoids is shown in Figure 3 separately for events predominantly constrained with cross-correlation data (median = 0.033 km, mean = 0.234 km) and for events constrained only with phase picks (median = 0.070 km, mean = 0.629 km). Note the long tails of the error distributions. Table 1 summarizes the uncertainty estimates for the individual tectonic regions. They are largest in the MTJ region, and relatively small in the three other tectonic areas.

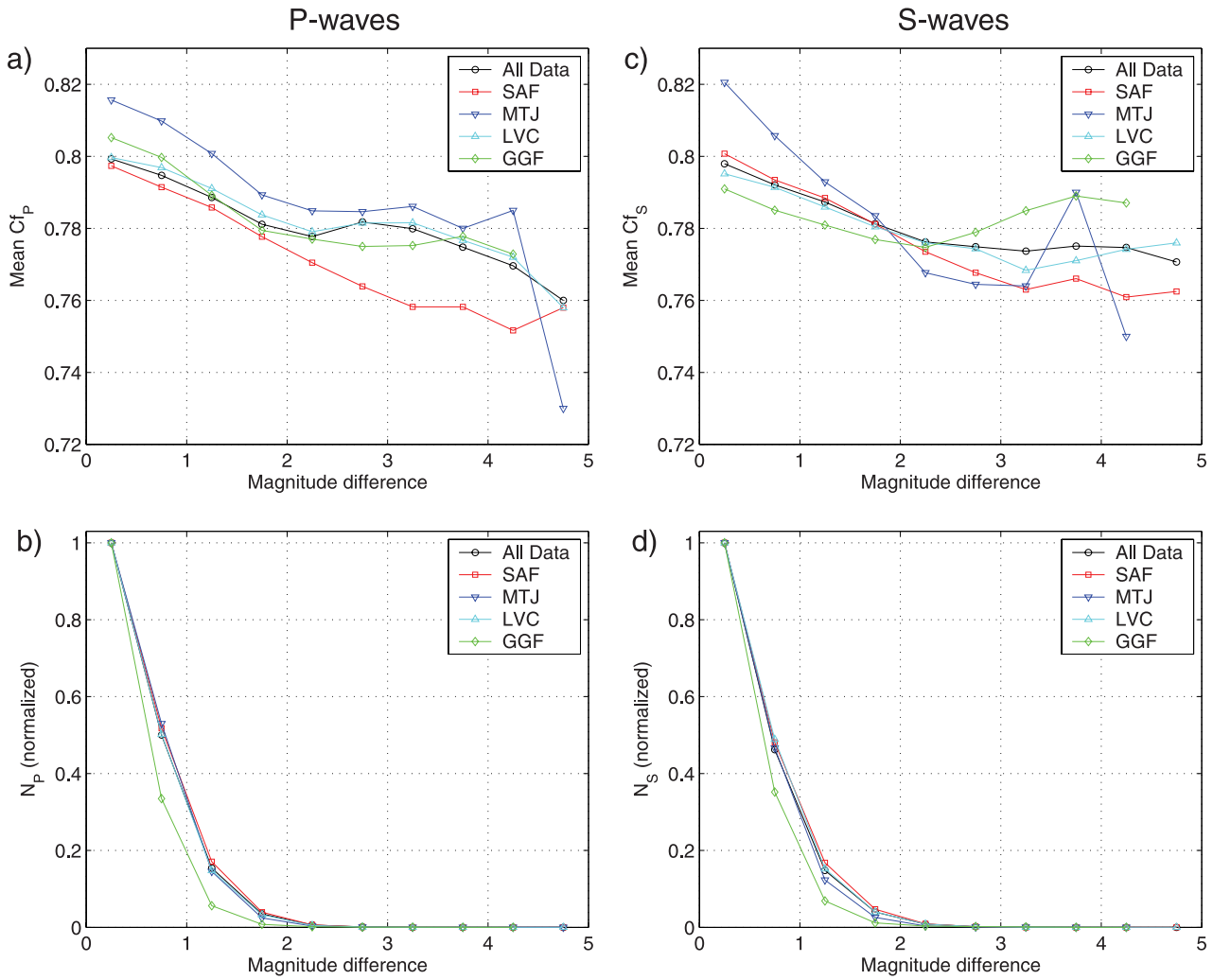

Figure 7. Same as Figure 6, but now the distributions of (a, c) *P* wave and *S* wave cross-correlation coefficients and (b, d) normalized number of correlations are shown as means within bins of 0.5 magnitude (M_L) difference.

[20] The relocated earthquakes reveal a focused view of the complex distribution of seismicity of Northern California (Figure 4, left), compared to the corresponding network locations (Figure 4, right). In particular, the new locations image previously hidden detailed networks of discrete faults at seismogenic depths that accommodate the stress imposed by the diverse tectonic forces associated with transform (SAF, MTJ), subduction (MTJ), and volcanic processes (LVC). Faults outlined by the relocated seismicity often correlate with the general trend of the fault lines mapped at the surface. At the Geysers Geothermal Field (GGF), fault orientations are more diverse and complex because of the nature of the underlying anthropogenic processes, thus leading to a more 'cloudy' image of the relocated seismicity, compared to the sharp images of near-vertical faults in the other areas.

[21] The double-difference catalog includes both correlated earthquakes that are located to the accuracy of the cross-correlation data and earthquakes that do not correlate that are located to the accuracy of the phase pick data. We find that 90% (or 185,601 events) of all earthquakes with digital waveforms available from the NCEDC correlate (Figure 5). We define an earthquake pair as correlated when at least four first-arriving *P* wave trains are similar at a cross-correlation coefficient of 0.7 or greater in the frequency band 1.5–15 Hz. Similar percentage values of 94% and 87% are found when we require at least 3 and 5 similar *P* waves trains, respectively. Correlated earthquakes occur widespread across Northern California with high concentrations being observed in a variety of tectonic settings that include predominantly strike-slip transform faults (SAF, MTJ), subduction zones (MTJ), volcanic areas (LVC), and geothermal fields (GGF) (Figure 5; Table 1).

[22] The decrease of the *P* wave correlation coefficients for correlated event pairs with increasing hypocentral separation is remarkably similar within the four regions, dropping from an average of Cf $\sim$ 0.9 for nearby hypocenters to $\sim$0.78 for events separated by 5 km (Figure 6a). The highest decay rate is observed for event pairs in the SAF region, where the abundance of streaks and repeating events along the San Andreas and Calaveras faults [*Rubin et al.*, 1999; *Schaff et al.*, 2002; *Waldhauser et al.*, 2004]

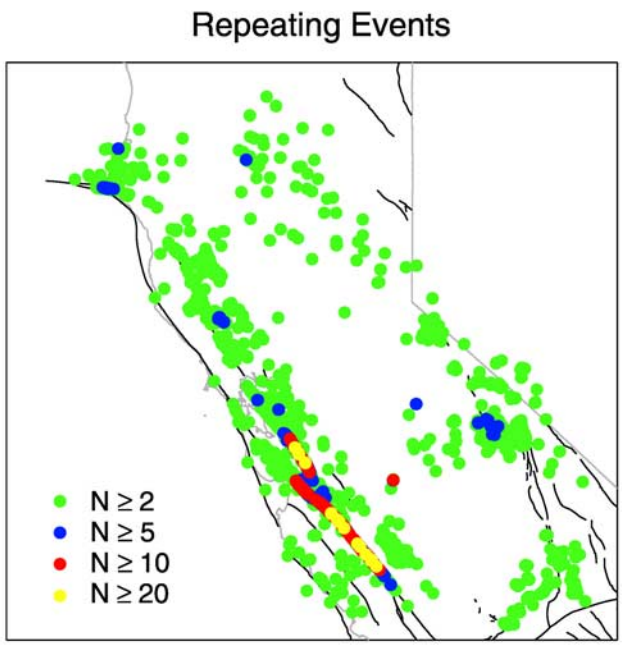

Figure 8. Distribution of clusters of repeating earthquakes in Northern California. N = number of events in each cluster. Gray lines denote the California state boundary, and the black lines mapped fault traces.

produces high Cf values at short separation distances, and the strong and complex structural variations caused by the interaction of the Pacific and American plate result in a rapid decay of Cf with increasing event separations. Coefficients are consistently lowest at the GGF where the events are induced by geothermal production activities, fracturing undisturbed porous rock along new faults whose orientations are random and varies rapidly over short distances [*Oppenheimer*, 1986]. In addition, GGF may exhibit time-dependent short-wavelength velocity variations because of the movement of fluids. As a result, the majority of earthquakes at GGF correlate over distances less than 1 km (Figure 6b). In comparison, correlation measurements in the three tectonic regions SAF, LVC, and MTJ are obtained over a broad range of separation distances between 0 and 5 km (Figure 6b).

[23] *S* waves show Cf values that are overall lower and break down faster with increasing hypocenter separation compared to the *P* wave coefficients because of their shorter wavelengths due to the slower wavespeed (Figure 6c). *S* waves typically correlate on pairs separated by less than 2 km, regardless of the type of tectonic region in which they occur (Figure 6d). *S* waves also tend to decorrelate faster than *P* waves because of their contamination with dissimilar *P* wave coda.

[24] The dependency of Cf on the difference in magnitudes (M_L) is shown in Figure 7. Correlation coefficients of both *P* and *S* waves decrease linearly with increasing difference between an event pairs' magnitudes (Figures 7a and 7c), with most correlations obtained for pairs with a magnitude differences less than 2 (Figures 7b and 7c).

[25] Our results show that the ability for two events to produce similar seismograms (in the frequency band 1.5–15 Hz), from which we can precisely measure phase delay times at common stations, primarily depends on the distance between their hypocenters and the difference between their magnitudes, and less so on the tectonic environment in which the events occur. This indicates that most of the seismicity in Northern California occurs along repeatedly breaking faults that are sufficiently smooth and long to generate earthquakes with similar seismograms over long separation distances.

4. Evaluation of Delay Time Measurement and Location Improvement

[26] The precision of the cross-correlation measurements and the improvement over existing picks is most readily assessed by using repeating earthquakes — a special category of correlated earthquakes that rupture the same fault patch more than once and therefore exhibit highly correlated waveforms and virtually zero delay times [*Poupinet et al.*, 1984; *Vidale et al.*, 1994; *Nadeau et al.*, 1995]. Repeating events have been shown to exist predominantly in the creeping sections of the San Andreas fault system. We search for repeating events in the double-difference catalog by selecting all pairs of events that produce *P* wave trains with a mean Cf $\geq$ 0.9 at 5 or more common stations, have well constrained hypocentral separations that are smaller than their respective rupture radius calculated from a circular, 3 MPa stress drop model, and have similar magnitudes ($\pm M_L 0.3$) [*Waldhauser and Ellsworth*, 2002]. We find a total of 24,438 repeating events that represent 12% of all events with waveforms. They occur in 7,406 clusters of between 2 and 33 events with magnitudes up to M_L 4.3 throughout Northern California (Figure 8; Table 1). While sequences with at least 2 repeating events are widespread, sequences with at least 5 events concentrate in the four regions SAF, MTJ, LVC, GGF (Figure 8). We find sequences with at least 10 events only on the creeping section of the San Andreas and Calaveras faults where they appear to image the transfer of fault creep at seismogenic depths from the San Andreas to the Calaveras fault.

[27] The median of the absolute cross-correlation differential times of these repeating events, after subtracting the mean in each cluster, is 0.002 s for both *P* and *S* waves, and the standard deviation (SD) is 0.01 s (Figure 9). In comparison, the median of the corresponding absolute differential times formed from the *P* wave picks is 0.023 s (SD = 0.15 s), which is $\sim$14 times less precise than the cross-correlation data for repeating events. These metrics are derived from the original measurements before relocation and therefore include the outliers that form the long tails of the distributions, especially in the cross-correlation data (see Figure 9). These outliers are easily detected by their large residuals and typically down-weighted or removed during the double-difference inversions. The precision of the delay time measurements decreases with increasing hypocenter separation, as waveforms become more dissimilar because of changes in the focal mechanisms and differences in the ray paths [e.g., *Waldhauser and Ellsworth*, 2000; *Schaff et al.*, 2004].

[28] Since we locate highly correlated repeating events to the precision of several meters to a few tens of meter [*Waldhauser and Ellsworth*, 2000; *Rubin*, 2002], the devia-

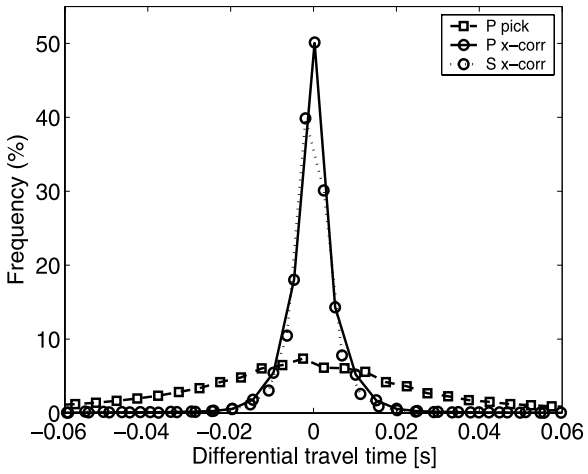

Figure 9. Histogram of *P* and *S* wave differential times of 24,438 repeating events measured via waveform cross-correlation (solid and dotted lines) and computed from NCSN phase picks (dashed line). The median of the cross-correlation absolute differential times is 0.002 s, and the standard deviation (SD) is 0.01 s. *P* wave pick delay times have a median of 0.023 s and an SD of 0.15 s.

tions of the corresponding network locations from the centroid location of each group of repeating events reflect their relative location error (Figure 10a). We find that these network locations have errors at the 95% confidence level of 0.7 km horizontally and 2 km vertically, and maximum mislocations of 8 km and 7 km, respectively (Table 1). Cross-correlation-based DD locations for repeating events in well-monitored regions thus represent a relative location improvement of up to a factor of ~1000 over existing network locations. The greater improvement in vertical control is due to the additional *S* wave differential times obtained via cross-correlation. Both network and double-difference relative location errors are largest for events near MTJ, and smallest for those in the GGF region (Table 1), reflecting differences in availability and coverage of seismic stations in the two regions.

[29] Double-difference locations of the repeating events based on phase picks alone (i.e., only minimizing model errors in the network locations but not reducing pick uncertainty) have errors at the 95% confidence level of 0.17 km horizontally and 0.7 km vertically (Figure 10b), indicating a factor of ~4 improvement in location precision over existing network locations. The significant improvement obtained by applying double-differences to picks alone is also visually demonstrated in Figure 4 (middle). Pick-based DD locations are closer to the CC-based DD locations than they are to the catalog locations, imaging detailed fault structures at the scale of a few hundreds of meters.

[30] One would expect the improvement in location precision to be greater for small events than for events with larger magnitudes, as cross-correlation can more easily improve on hard to pick phase onsets for small events and measure additional differential times on seismograms not picked because of low signal to noise ratio. In Figure 11 we show the lateral and vertical deviation of the locations of repeating events from their respective cluster centroid as a function of event magnitude. We observe a slight deviation increase with increasing magnitudes for CC-based DD locations (Figure 11, circles), which reflects the way we determine the hypocenter separation cutoff in our search procedure for repeating events (i.e., as a function of estimated rupture dimension, assuming a 30 bar constant stress drop model; see gray line in Figure 11). Epicentral deviations from the centroid of pick-based DD solutions (Figure 11, squares) are larger but increase similarly to the CC-based solutions, while depth deviations decrease from about 0.8 km for small events to 0.5 km for events with M ~ 3.5 events. Surprisingly, deviations from the centroid increase with increasing magnitudes for network locations (Figure 11, diamonds), indicating that both pick- and CC-based DD solutions appear to produce the greatest location improvements over network locations for larger magnitude

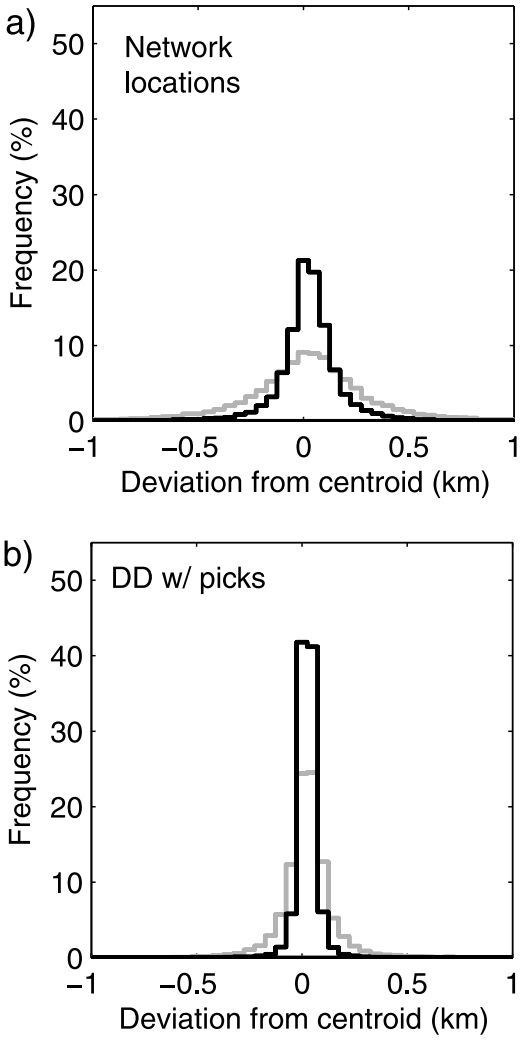

Figure 10. Histograms of horizontal (black) and vertical (gray) locations of 24,438 repeating earthquakes computed relative to the centroid of their respective cluster. Shown are (a) network locations and (b) double-difference solutions based on pick data.

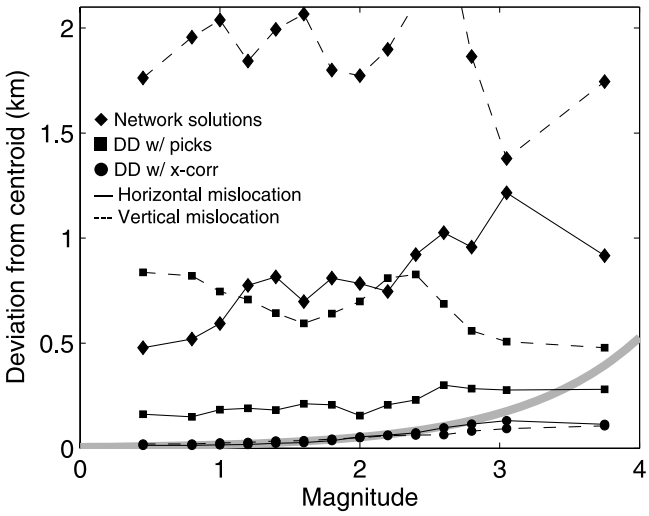

Figure 11. Distribution of horizontal (solid line) and vertical (dashed line) deviations of the location of repeating events from their respective cluster centroid as a function of magnitude. Shown are network locations and DD locations based on pick data and cross-correlation data. Deviations at the 95% confidence level are shown for bins of 0.2-magnitude intervals. Gray line indicates approximate rupture dimension computed from a circular 3-MPa stress drop model (see text for explanation).

repeating events. The reason is probably because the 1-D models used to locate events at the NCSN break down with distance, so that more distant picks are not well predicted by these models. Since these picks are still fully weighted in the locations, they actually degrade the location quality slightly compared to smaller events with picks from closer stations.

[31] A more general assessment of differences between network and relocated locations for larger magnitude events is presented in Figure 12a. The figure shows epicentral shifts between events with $M \geq 4.0$ in the DD catalog and their respective location in the NCSN catalog (as per September 2006). Most of the larger shifts occur at the edge of the network in areas with bad station coverage, such as near MTJ and the eastern border of California. A few significantly large differences, however, occur in well-monitored regions such as the central SAF system. We investigate the largest of these differences, associated with two aftershocks (M4.2 and M4.8) of the M7.0 Loma Prieta earthquake that both occurred on 18 October 1989 within 3 hours of the mainshock. Network locations place the two aftershocks (NCSN IDs 10090186 and 10090486) about 2 km south of the mainshock, which occurred ~8 km southwest of and perpendicular to the main surface trace of the San Andreas Fault (Figure 12b). The DD locations of the same aftershocks are near the surface trace of the San Andreas fault, about 7 km to the east of the network locations. Their RMS values are 0.007 and 0.009 s, respectively. Inspection of the differential time links indicates that a significant number of phase arrival times reported in the NCSN catalog appeared to be misidentified or misassociated and were removed during the DD relocation process. Misidentifications of phase arrival times are easily possible for aftershocks of large earthquakes as they are often embedded in the coda of prior aftershocks as is the case here, and can cause significant location bias when used in single-event location procedures as employed at the NCSN. The questionable quality of these two aftershock locations in the NCSN catalog has also been expressed by their reported high RMS values of 0.87 and 0.75 s. (Note that since our reanalysis the network locations for these two aftershocks have been

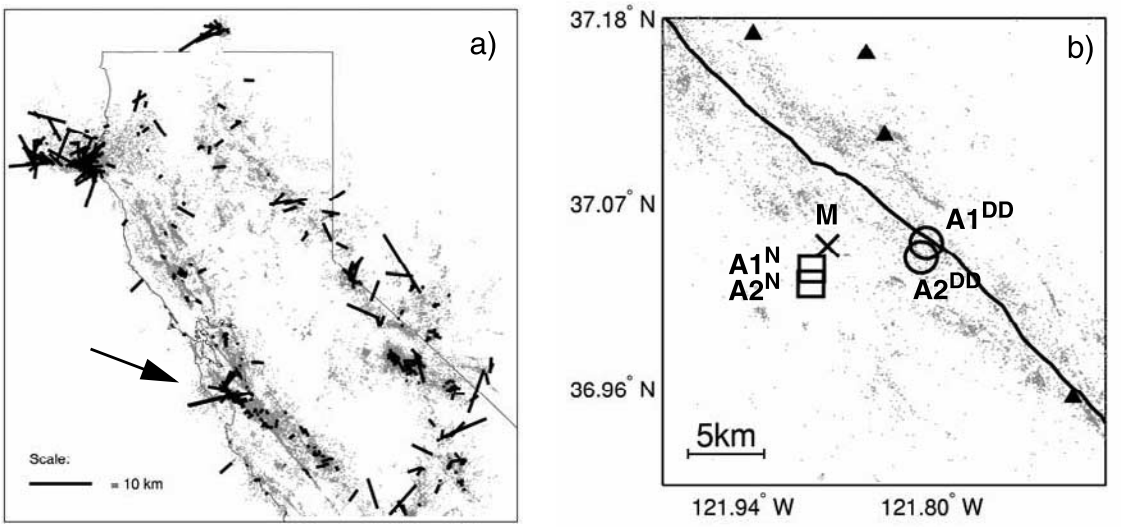

Figure 12. (a) Differences between epicenter locations in the NCSN catalog and those in the DD catalog for events with $M \geq 4.0$. Solid black lines indicate azimuth and scaled distance between the two locations, and thin lines denote state boundary. Arrow points to area shown in Figure 12b. (b) Two aftershocks (A1 and A2) of the Loma Prieta mainshock (M) shown at their network locations (squares) and at their relocated DD location (circles). Solid line denotes surface exposure of the SAF, and gray dots denote DD located seismicity.

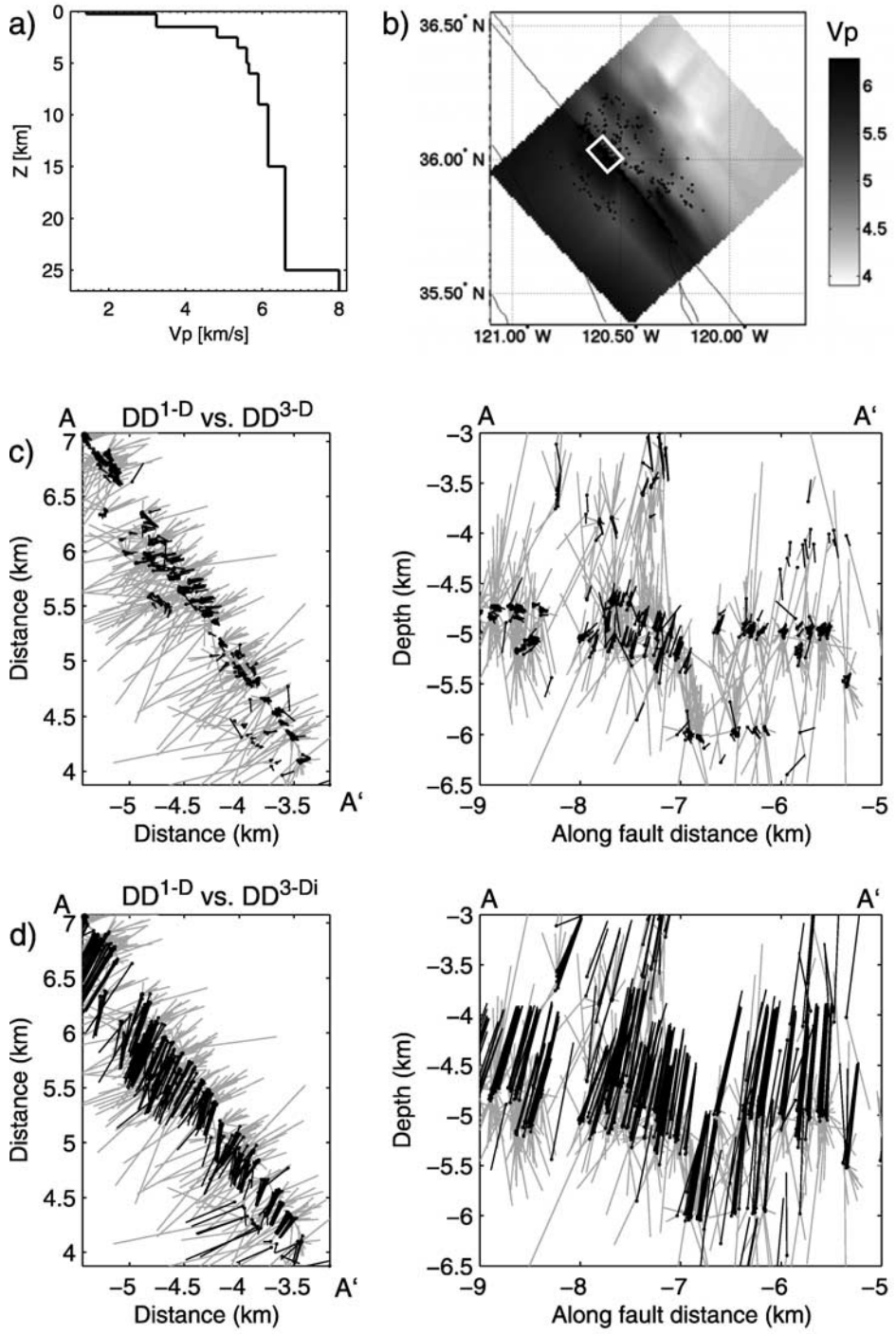

Figure 13. Location bias in 1-D double-difference solutions due to unmodeled 3-D effects. (a) Velocity-Depth function used to compute the DD catalog locations in the Parkfield area. (b) 3-D high-resolution tomographic model of the Parkfield area [from *Thurber et al.*, 2006]. Horizontal cross-section (taken at 4 km depth) shows the strong velocity contrast across the fault. Solid dots denote seismicity investigated. White box includes events shown in Figures 13c and 13d. (c, left) Map view and (right) longitudinal cross-section showing $DD^{1\text{-}D}$ locations (i.e., the DD catalog locations; dots), with black lines connecting to the corresponding $DD^{3\text{-}D}$ locations that are relocated in a 3-D model [*Thurber et al.*, 2006]. In both 1-D and 3-D applications, starting locations were taken as the NCSN network locations. A selected area (see box in Figure 13b) of the relocated data set is depicted to show details. (d) Same as in Figure 13c but with the black lines connecting the $DD^{1\text{-}D}$ locations (dots) with the corresponding $DD^{3\text{-}Di}$ locations that are both located and relocated in a 3-D model [*Thurber et al.*, 2006]. Gray lines connect $DD^{1\text{-}D}$ location to NCSN locations for comparison. Note the systematic shifts in Figures 13c and 13d that indicate small differences in relative locations (see text for details and Table 2 for statistics).

Table 2. Median Differences (in Meters) in Relative (Δdx, Δdy, Δdz) and Absolute (Δx, Δy, Δz) Locations Between Solutions Computed With Different Models and Starting Locations for 2213 Well-Recorded Earthquakes Along the San Andreas Fault near Parkfield[a]

	Relative Location Differences			Absolute Location Differences		
	Δdx	Δdy	Δdz	Δx	Δy	Δz
$DD^{1\text{-}D} - DD^{3\text{-}D}$	7	7	20	51	32	115
$DD^{1\text{-}D} - DD^{3\text{-}Di}$	7	6	17	170	299	877
$NCSN - DD^{3\text{-}Di}$	228	169	388	282	349	952

[a]$DD^{1\text{-}D}$ are double-difference relocations computed using a layered 1-D model (Figure 13a), with starting locations taken from the NCSN catalog (i.e., 1-D single-event locations). $DD^{3\text{-}D}$ are double-difference relocations computed in the 3-D model of *Thurber et al.* [2006] (Figure 13b), starting from NCSN catalog locations; $DD^{3\text{-}Di}$ are the same as $DD^{3\text{-}D}$ but starting from single-event locations obtained in the 3-D model [*Thurber et al.*, 2006]. Relative distances between events (dx, dy, dz) are calculated between each event in the DD catalog and its 10 nearest neighbors and compared with the corresponding relative distances in the two other sets of locations.

revised at the NCSN, and are now near the DD locations with RMS values of 0.09 s and 0.07 s.)

5. Discussion and Conclusions

[32] The double-difference relocations presented in this study indicate a significant improvement in location precision over the existing single-event locations determined on a routine basis and listed in the NCSN catalog. The relative location improvement is in part due to the improvement in delay time measurement using cross-correlation, and in part due to the reduction of model errors in the NCSN locations using double-differences. Since the latter is carried out by a linearized inversion that requires the prediction of the observed data, the difference between the model used to predict the data and the true structure may bias the DD locations. In this study we have relied on well-established 1-D (depth-dependent) velocity models (see, for example, Figure 13a) to solve the forward problem. These models are able to predict the observed data very efficiently, a crucial aspect in an application of the scale presented here. Furthermore, they provide the most consistent representation of the crustal structure of Northern California, although more detailed 3-D structural information from passive and active source tomographic investigations are available for selected regions [e.g., *Thurber et al.*, 2006; *Hardebeck et al.*, 2007].

[33] To estimate potential location bias due to unmodeled 3-D structures in our DD solutions based on 1-D models we compute double-difference solutions for 4332 well-recorded events on the San Andreas fault near Parkfield in the high-resolution 3-D *P* wave velocity model of *Thurber et al.* [2006] ($DD^{3\text{-}D}$). The 3-D velocity structure in the Parkfield area is complex, with a significant velocity contrast across the near-vertical fault (Figure 13b). The corresponding *S*-velocity model used to predict the *S* wave differential times is obtained by scaling the *P* model by a factor of 1.73. Figure 13a shows the NCSN 1-D model used to compute the DD catalog locations in the Parkfield area ($DD^{1\text{-}D}$). A comparison between the 1-D and 3-D DD locations of events near the fault indicates that the changes are systematic and thus differences in relative locations small (Figure 13c). The median differences in relative locations between the two data sets are 7 m in both horizontal directions and 20 m vertically (Table 2). These statistics are based on relative locations computed between each event and its ten nearest neighbors in the $DD^{1\text{-}D}$ catalog, and the corresponding relative locations in the $DD^{3\text{-}D}$ catalog. Median differences in absolute locations are 51 m in east–west, 32 m in north–south, and 115 m in vertical directions (Table 2).

[34] Although the linearized double-difference equations would appear to produce only relative locations, experiments with synthetic data have shown that the iterated solutions converge toward the true absolute locations despite gross differences between the velocity models used to create and model the data [*Waldhauser and Ellsworth*, 2000; *Menke and Schaff*, 2004]. The degree to which such corrections can be resolved, however, depends primarily on the quality and distribution of the data. To ensure robustness during catalog relocation in areas with sparse station coverage we have typically damped changes in the location of cluster centroids. Thus potential systematic bias in the absolute network locations are not accounted for in the relocated DD catalog presented here. In the Parkfield area, for example, single-event locations computed in a high-resolution 3-D model and subsequently relocated using double-differences together with the same 3-D model ($DD^{3\text{-}Di}$; *Thurber et al.* [2006]) locate on average ~449 m southwest of and 952 m shallower than the NCSN locations (Table 2; gray lines in Figure 13d). This is because the single-event NCSN locations are based on a 1-D model (Figure 13a) and therefore do not account for the sharp velocity contrast across the San Andreas Fault (Figure 13b). Consequently, the $DD^{1\text{-}D}$ locations inherit that systematic bias since they are relocated starting from the NCSN locations (black lines in Figure 13a). Nevertheless, their absolute locations are slightly better compared to the NCSN locations because of the improvement in relative locations, with median differences of less than 300 m in horizontal and 877 m in vertical directions. The differences in relative locations between the $DD^{1\text{-}D}$ solutions and the $DD^{3\text{-}Di}$ solutions are again small (median = 7 m laterally and 17 m vertically; see Table 2).

[35] These comparisons demonstrate that the relative hypocenter locations in the DD catalog are relatively robust against deviations of a reasonable 1-D model from the highly heterogeneous structures resolved by 3-D tomography, and against differences in starting locations taken from the NCSN catalog and those determined in a 3-D model. Although systematic location bias in the NCSN starting locations are not corrected for in the new catalog, absolute DD locations can still be better than those of the corresponding network solutions. This is particularly true in areas with little or no systematic bias in the centroid of network locations that form a continuously linked cluster, but less so in areas where complex velocity structure and/or sparse station coverage may introduce such bias. Thus

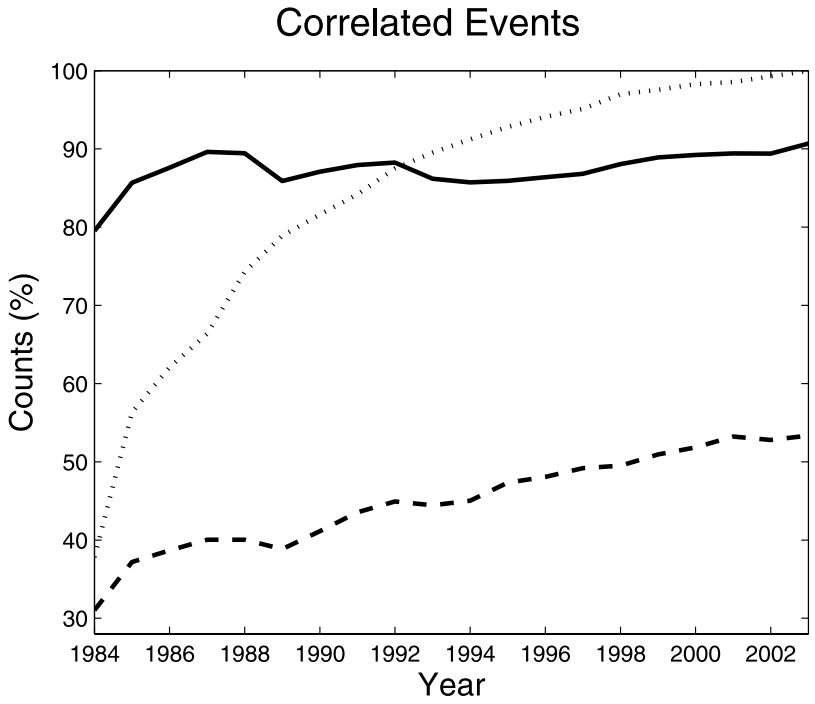

Figure 14. Increase in the number of correlated events since 1984, the start of the digital NCSN archive. Percentages are shown for the total number of correlated events between 1984 and each subsequent year (solid line) and as the mean of percentages of correlated events in each cell of 20 × 20 km (dashed; see Figure 5 for cell locations). Dotted line indicates the relative increase in the number of cells shown in Figure 5 (i.e., the approximate increase in area covered by seismicity).

future updates of the DD catalog may benefit from using regional 3-D velocity models for Northern California [e.g., *Thurber et al.*, 2007; *Hardebeck et al.*, 2007] for determining both accurate single-event absolute locations and subsequently precise double-difference relative locations. Then, the sharp structures of seismicity observed at depth can be more reliably correlated with geologic observations made at the surface, such as mapped surface fault traces.

[36] The study presented here shows that hypocentral separation, and thus event density, is one of the primary controlling factors in improving existing single-event locations by means of cross-correlation and double-difference methods. Seismic archives must therefore be given the time to consistently grow and accumulate a critical number of events in any given area. Between 1984, the year digital recording and archiving began at the NCSS, and 2003 the percentage of correlated events increased from 80% during 1984 to 90% during 1984–2003 (solid line in Figure 14). We note that these percentages are dominated by the locally concentrated occurrence of repeating earthquakes along the creeping San Andreas faults, and by the dense distribution of seismicity at GGF and LVC. If we account for the varying seismicity rates across Northern California by computing the average of the percentages of correlated events in each cell as a function of time (see Figure 5 for cell locations), then these percentages increase linearly from 30% for events during 1984 to 54% during 1984–2003 (dashed line in Figure 14). The linear increase demonstrates that the NCSN and NCEDC's consistent long-term seismic monitoring practices and data archiving policies will lead to continued improvement in the location of events that occurred in the past as well as new events as this archive continues to grow. Periodic cross-correlation-based double-difference reanalysis of this data archive may thus become part of the routine network procedure.

[37] With archives of digital seismic data growing around the world because of the continued need for monitoring earthquake activity and compliance with the nuclear test ban treaty, a reanalysis of these archives following the procedures described in this study is expected to improve the location precision in existing catalogs. Furthermore, implementation of cross-correlation and double-difference methods into routine location procedures can produce highly accurate relative locations of new events in near real time. High-resolution catalogs of past seismicity as well as the immediate knowledge of the precise location of a new event relative to the background seismicity are of broad significance in the scientific study of earthquakes and Earth structure, and have considerable social and economic impact in the evaluation and mitigation of seismic hazard.

[38] The double-difference catalog described in this paper is available from the authors on request.

[39] **Acknowledgments.** We are grateful to the following institutions that contributed with their seismic networks to the NCSS data used in this study: U.S. Geological Survey, Menlo Park; University of California, Berkeley; California Institute of Technology; University of Nevada, Reno; California Division of Water Resources; University of Utah; and University of Southern California. We thank the NCSN personnel from the U.S. Geological Survey and the NCEDC personnel from the Berkeley Seismological Laboratory for their gargantuan effort in building and maintaining the high-quality seismic archive and making it easily accessible and the LDEO computer support group for their assistance in computational matters. We appreciate valuable comments and suggestions by Associate Editor John Townend, Jeanne Hardebeck, Jean-Luc Got, and David Oppenheimer that helped improve the manuscript. This research was supported by USGS-NEHRP grants 05HQGR0051 and 06HQGR0054,

with additional support from the Southern California Earthquake Center (SCEC grant 076547). This is LDEO contribution number 7154.

References

Deichmann, N., and M. Garcia-Fernandez (1992), Rupture geometry from high-precision relative hypocenter locations of microearthquake clusters, *Geophys. J. Int.*, *110*, 501–517.

Geiger, L. (1910), Herdbestimmung bei Erdbeben aus den Ankunftszeiten, *K. Ges. Wiss. Gött.*, *4*, 331–349.

Got, J.-L., and P. Okubo (2003), New insights into Kilauea's volcano dynamics brought by large-scale relocation of microearthquakes, *J. Geophys. Res.*, *108*(B7), 2337, doi:10.1029/2002JB002060.

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., A. J. Michael, and T. M. Brocher (2007), Seismic velocity structure and seismotectonics of the eastern San Francisco Bay Region, California, *Bull. Seismol. Soc. Am.*, *97*, 826–842, doi:10.1785/0120060032.

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.

Klein, F. W. (2002), User's guide to HYPOINVERSE2000, a Fortran program to solve for earthquake locations and magnitudes, *U.S. Geol. Surv. Open-File Rep., 02-172*, 123 pp., 01-113, Menlo Park, California.

Menke, W., and D. Schaff (2004), Absolute earthquake locations with differential data, *Bull. Seismol. Soc. Am.*, *94*, 2254–2264, doi:10.1785/0120040033.

Nadeau, R. M., W. Foxall, and T. V. McEvilly (1995), Clustering and periodic recurrence of microearthquakes on the San Andreas Fault at Parkfield, *Science*, *267*, 503–507.

Oppenheimer, D. (1986), Extentional tectonics at the Geysers Geothermal Area, California, *J. Geophys. Res.*, *91*, 11,463.

Oppenheimer, D. H., F. W. Klein, J. P. Eaton, and F. W. Lester (1993), The Northern California Seismic Network bulletin, January–December 1992, *U.S. Geol. Surv. Open-File Rep., 93-578*, Menlo Park, Calif.

Paige, C. C., and M. A. Saunders (1982), LSQR: Sparse linear equations and least squares problems, *ACM Trans. Math. Software*, *8/2*, 195–209.

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.

Prejean, St., W. L. Ellsworth, M. Zoback, and F. Waldhauser (2002), Fault structure and kinematics of the Long Valley Caldera region, California, revealed by high-accuracy earthquake hypocenters and focal mechanism stress inversions, *J. Geophys. Res.*, *107*(A11), 1397, doi:10.1029/2001JB001168.

Richards, P. G., F. Waldhauser, D. P. Schaff, and W.-Y. Kim (2006), The applicability of modern methods of earthquake location, *Pure Appl. Geophys.*, *163*, 351–372.

Rubin, A. (2002), Using repeating earthquakes to correct high-precision earthquake catalogs for time-dependent station delays, *Bull. Seismol. Soc. Am.*, *92*, 1647–1659.

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.

Schaff, D. P., G. Bokelmann, G. C. Beroza, F. Waldhauser, and W. L. Ellsworth (2002), High-resolution image of Calaveras Fault seismicity, *J. Geophys. Res.*, *107*(B9), 2186, doi:10.1029/2001JB000633.

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

Shearer, P., 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., H. Zhang, F. Waldhauser, J. Hardebeck, A. Michael, and D. Eberhart-Phillips (2006), Three-dimensional compressional wavespeed model, earthquake relocations, and focal mechanisms for the Parkfield, California, region, *Bull. Seismol. Soc. Am.*, *96*, 38–49.

Thurber, C., T. Brocher, H. Zhang, and V. Langenheim (2007), Three-dimensional *P*-wave velocity model for the San Francisco Bay region, California, *J. Geophys. Res.*, *112*, B07313, doi:10.1029/2006JB004682.

Vidale, J. E., W. L. Ellsworth, A. Cole, and C. Marone (1994), Variations in rupture process with recurrence interval in a repeated small earthquake, *Nature*, *368*, 624–626.

Waldhauser, F. (2001), HypoDD: A computer program to compute double-difference earthquake locations, *U.S. Geol. Surv. Open-File Rep., 01-113*, Menlo Park, California.

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

Waldhauser, F., and W. L. Ellsworth (2002), Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations, *J. Geophys. Res.*, *107*(B3), 2054, doi:10.1029/2000JB000084.

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.

Waldhauser, F., W. L. Ellsworth, D. P. Schaff, and A. Cole (2004), Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett.*, *31*, L18608, doi:10.1029/2004GL020649.

D. P. Schaff and F. Waldhauser, Lamont-Doherty Earth Observatory, Columbia University, P.O. Box 1000, 61 Route 9W, Palisades, NY 10964, USA. (felixw@ldeo.columbia.edu)