Final Technical Report

USGS/NEHRP Award No: 05HQGR0051

Improving Earthquake Locations in Northern California Using Waveform Based Differential Time Measurements

Principal Investigator: Felix Waldhauser
Co-PI: David Schaff

Lamont-Doherty Earth Observatory of Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

Project Period: 01/01/2005 – 12/31/2006

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number 05HQGR0051. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

ABSTRACT

We relocated the earthquake catalog of the Northern California Seismic Network (NCSN), using a recently computed comprehensive database of cross-correlation differential times for earthquakes between 1984-2003. The correlation data base includes approximately 3 billion P- and S-wave differential times that are calculated from pairs of waveforms recorded at common stations with cross correlation coefficients of 0.7 or greater, for events separated by less than 5 km. The cross-correlation data is combined with differential times computed from NCSN phase picks to relocate the hypocenters with the double-difference algorithm. All event pairs with at least 6 pair-wise observed phases are relocated (311,273 events). The relocated catalog has a root mean square (RMS) of the weighted pick based differential-time residuals of 0.017 s, compared to 0.124 s before relocation. The weighted RMS of the cross-correlation data is 0.004 s after relocation. Location precisions range from tens to hundreds of meter. The double-difference locations reveal a focused view of the complex distribution of seismicity in northern California. Discrete faults are imaged with unprecedented detail in tectonic regions as diverse as the San Andreas Fault system (SAF), the Mendocino Triple Junction (MTJ), the Long Valley Caldera (LVC), and the region of induced earthquakes at the Geysers Geothermal Field (GGF). We find that 90% of the earthquakes have correlated P- and S-wave trains at common stations. The occurrence of correlated earthquakes is widespread across northern California with some of the highest concentrations being observed along the transform faults within the SAF system, in the LVC region, at the GGF, and at the MTJ, indicating the general applicability of these methods to improve hypocenter locations. Using repeating earthquakes we find that, at the 95% confidence level, network locations are mislocated on average by 0.7 km horizontally, and 2 km vertically. This study shows that consistent long-term seismic monitoring practices and data archiving policies, as is the case at the Northern California Seismic Network, are key to improve catalogued hypocenters obtained from standard single-event location methods using phase picks, and to determine accurate location of future events within the framework of earthquake monitoring.

1. Overview

This final report covers the activities performed between January 1, 2005 (start date of the project) and December 31, 2006. The work described in this report is being undertaken by the principle investigator Felix Waldhauser and by co-PI David Schaff. The research includes the quality assessment of a recently developed comprehensive waveform cross-correlation database and its use to relocate 311,273 well recorded earthquakes in the Northern California Seismic Network (NCSN) earthquake catalog between 1984-2003 using the double-difference method.

2. Investigations undertaken

Assessment of cross correlation data accuracy

We have assessed the quality of a recently computed database of cross correlation delay time measurements for Northern California (USGS/NEHRP grant 03HQGR0004; Schaff and Waldhauser, 2005) (Figure 1). The measurements are based on the complete waveform database (up to May 2003) stored at the Northern California Earthquake Data Center (NCEDC) at the University of Berkeley. 23 billion cross correlations were performed on pairs of seismograms recorded at common stations from pairs of events that are separated by less than 5 km. 1.7 billion P-wave and 1.2 billion S-wave differential times were obtained from pairs of seismograms that had cross correlation coefficients of 0.7 or larger. In addition to the cross correlation data we have computed approximately 1 billion travel time differences from P-phase picks listed in the earthquake bulletin of the Northern California Seismic Network (NCSN).

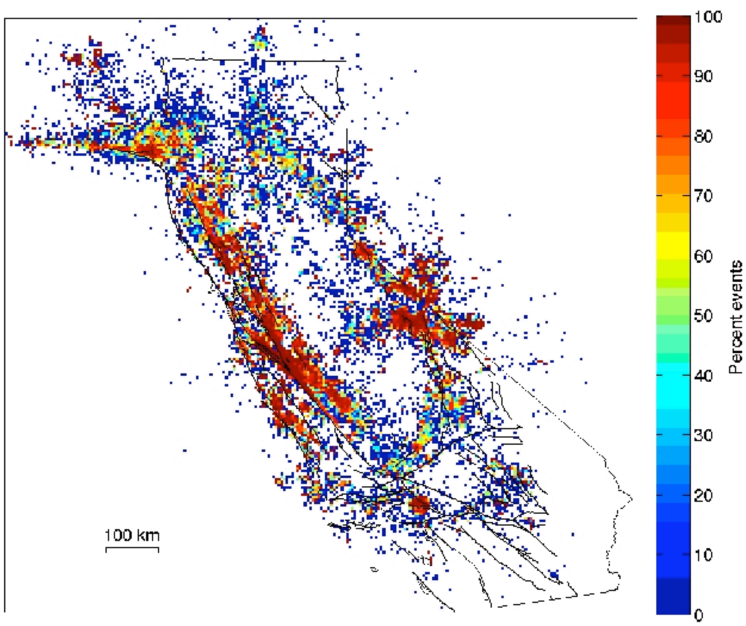

Figure 1 Percentage of correlated events that have cross-correlation coefficients of CC ≥0.7 with at least one other event recorded at four or more stations. Percentages are computed from the total number of events within bins of 5x5 km. (From Schaff and Waldhauser, 2005).

We compared our correlation differential times with cross correlation measurements made independently by Peter Shearer (pers. comm.) for earthquakes recorded at 470 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). Histograms of the differences between the two sets of differential times are shown in Figure 2 for events common in both data sets. From a total of 32,475 differential times compared, 96% of the P-wave data agree within 10 msec, and 63% within 1 msec (Figure 2a). 92% of the S-wave data agree within 10 msec, 59% within 1 msec (Figure 2b). Most differences between the two sets are smaller than the sampling rate (100 Hz). 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. Prior to using our data for relocation we detect outliers by comparing time measurements made for both 1 and 2 sec windows. Measurements with differences larger than one sample are typically removed. Correlated noise, for example, can be easily detected this way.

Histograms of differences in cross correlation coefficients are shown in Figure 2a and b for P-waves and S-waves respectively. Cross-correlation coefficients tend to be systematically higher for our measurements compared to those by Shearer. This is likely due to the different window lengths used by the two groups. Even though we computed both 1 and 2 sec windows for the NCSN data, only cross correlations over 1 sec window lengths are analyzed here. We typically use the 1 sec window length data for relocation purposes. Shearer computes the cross correlation function for window lengths of 2 and 3 seconds around the P-wave and S-wave train, respectively.

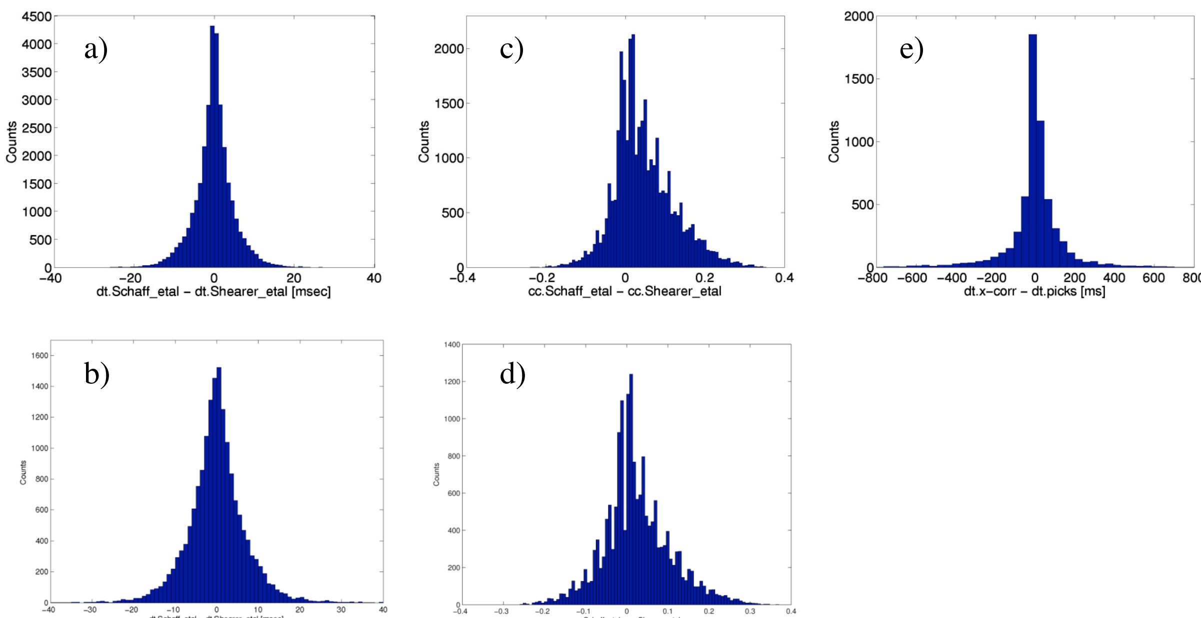

Figure 2 Histograms of differences between cross correlation measurements computed by Schaff and Waldhauser (2005) and P. Shearer (pers. comm.) for earthquakes near Mendocino, CA. a) P-delay times. b) S-delay times. c) P-wave cross correlation coefficients. d) S-wave coefficients. e) Differences between our cross correlation P-wave delay times and delay times formed from the NCSN phase picks.

A comparison of our cross correlation data with the corresponding travel-time differences formed from NCSN phase picks (Figure 2e) indicates a standard deviation of 150 msec, which is similar to the assumed pick accuracy for P-phases. S-phases are not routinely picked at the NCSN.

Catalog Relocation

The NCSN archive between January 1, 1984 and May 30, 2003 includes 408,084 events, nearly 7 million P-wave arrival times picked at ~900 stations, and 15 million digital waveform. We chose all 311,273 earthquakes recorded at 6 or more stations for relocation, using a combination of differential travel times computed from the NCSN P-wave picks and the 1.7 billion P-wave and 1.2 billion S-wave cross correlation differential times described in Schaff and Waldhauser (2005). Both pick and cross correlation differential times are combined in the 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 to the accuracy of the pick data (Waldhauser and Ellsworth, 2000).

We adapted *hypoDD* (Waldhauser, 2001) to run on a 32 node (64 CPUs) Linux cluster at Lamont to handle the vast amount of data. The program now runs in 'black box' mode in which the best relocation parameters are found automatically. The relocation problem has been regionalized to distribute the computational load to individual processors. We generated 513 overlapping boxes, each box including events that are connected to neighboring events through a web of differential time links not exceeding 3 million measurements. Double-difference solutions in 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 its cluster centroid.

We use 1D layered velocity models to predict travel times and partial derivatives. The models are chosen for each box from a compilation of 39 local 1D layered velocity models used by the NCSN to locate the earthquakes on a routine basis. Most of these models were determined in separate studies by a 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 functions to generate 28 layers of constant velocity to avoid strong velocity jumps across interfaces.

We investigated the effect that the 3D velocity structure has on 1D double-difference solutions. We used the Parkfield differential-time data set described in Waldhauser et al. (2004) together with a recently derived 3D model for the Parkfield area (Thurber et al., 2006) to obtain 3D *hypoDD* solutions and compared them to DD solutions obtained with a 1D layered velocity model (essentially those described in Waldhauser et al. 2004). Figure 3 shows the difference between the two results, in map view and cross section. The mean in absolute horizontal shifts between the two sets of locations is 140 m, and 180 m in vertical direction. The mean in relative location differences between each event and its neighbors within 2 km is only 33 m. Note that some of the discrepancies (red lines) appear to be caused by numerical instabilities in the 3D ray tracing routine, and further investigations are necessary to solve these problems. These results indicate that good 1D velocity models are a very efficient way to relocate earthquakes even in areas as structurally complex as the Parkfield section of the San Andreas fault.

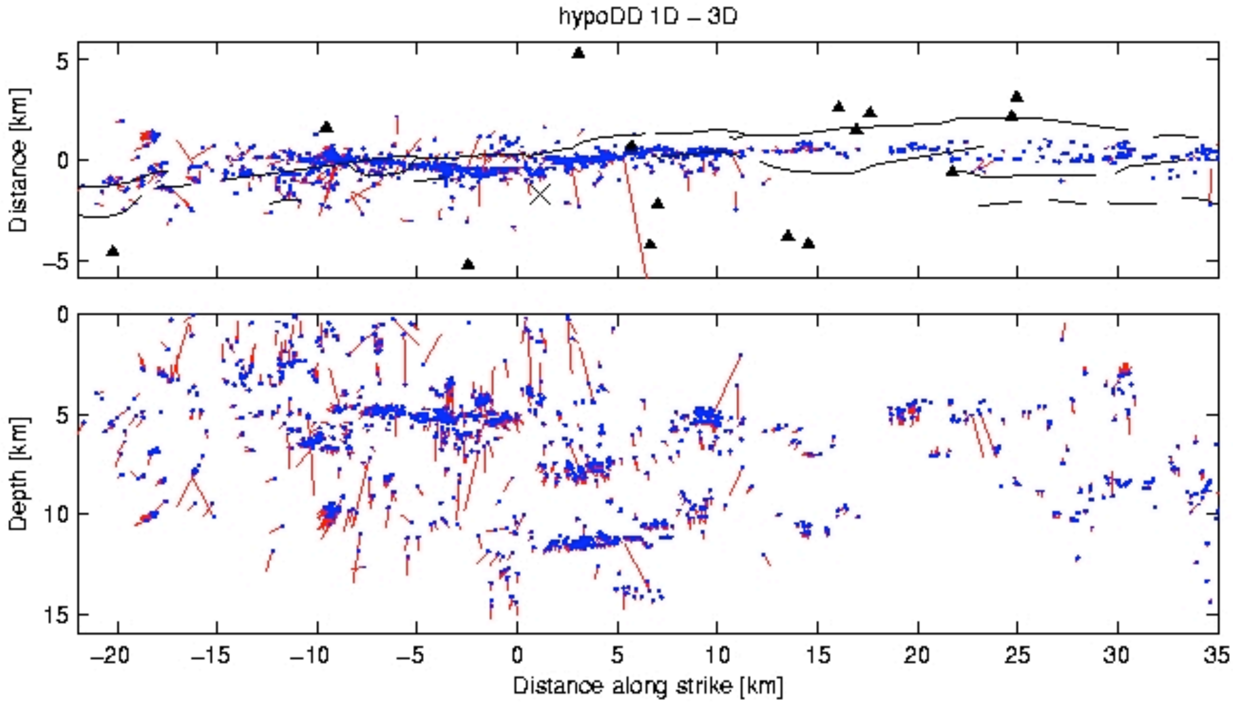

Figure 3 Comparison between 1D and 3D double-difference solutions. Top panel: Map view. Bottom panel: NW-SE cross-section. Blue dots: hypoDD locations obtained in a 1D layered velocity model (Waldhauser et al. 2004). Red lines point to the corresponding hypoDD locations obtained by ray tracing in a 3D velocity model.

2. Results

The resulting cross-correlation based double-difference catalog for Northern California currently exists as a beta version (NCAeq.DDcc.1.beta) and includes 311,273 events, or 98% of the initial number of earthquakes that went into the relocation process. Events are 'lost' during the iterative relocation process mostly due to insufficient data links after the weighting function removed outliers. The relocated catalog has a root mean square (RMS) of the weighted pick differential time residuals of 0.017 s, compared to 0.124 s before relocation. The weighted RMS of the cross-correlation data is 0.004 s after relocation. The new double-difference catalog reveals a focused view of the complex distribution of seismicity of northern California. Discrete faults are imaged with unprecedented detail in tectonic regions as diverse as the San Andreas Fault system (SAF), the Mendocino Triple Junction (MTJ), the Long Valley Caldera (LVC), and the region of induced earthquakes at the Geysers Geothermal Field (GGF) (Figure 4). In addition, artifacts in the NCSN locations (arrow in Figure 4), most likely caused by effects related to transitions in regional 1D models used for routine locations, are removed.

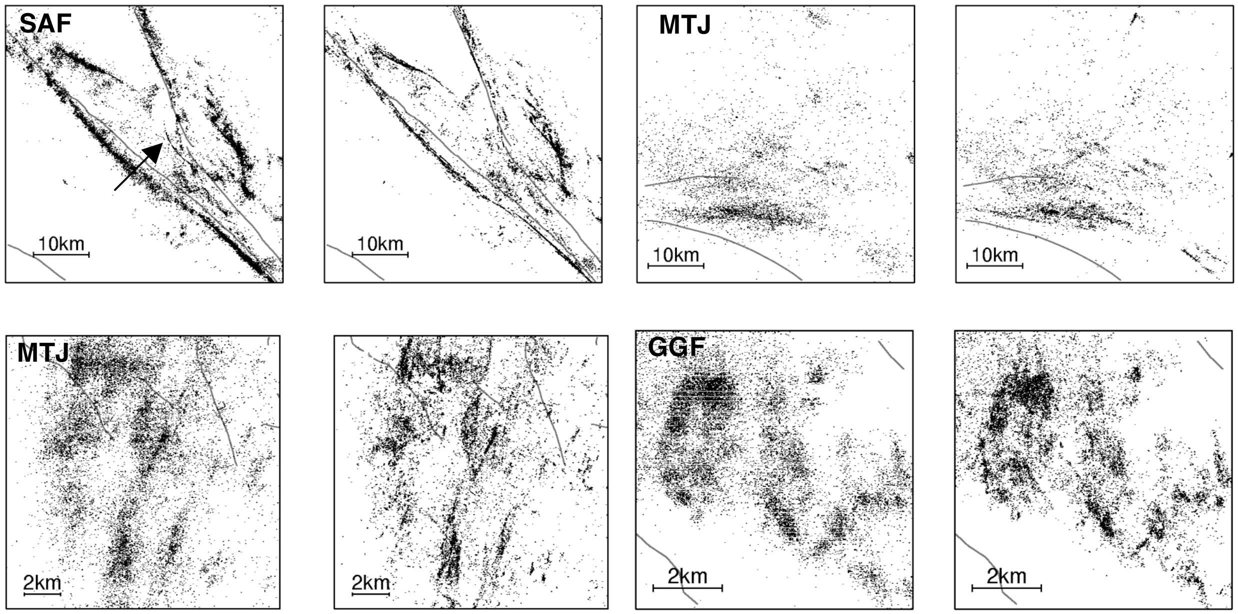

Figure 4 Comparison between network and double-difference locations. Map views of representative areas in each of the four focus regions are shown. Note the networks of discreet faults imaged in the tectonic regions SAF, MTJ, and LVC, compared to the ‘clouds’ imaged in the region of induced seismicity at GGF. Lines indicate traces of faults mapped at the surface.

We find that 90% (or 185,601 events) of all earthquakes with digital waveforms available from the NCEDC correlate. We define two earthquakes as correlated when at least four first-arriving P-wave trains in the frequency band 1.5-15 Hz are similar at a cross-correlation coefficient (CC) of 0.7 or greater. Similar percentage values of 94% and 87% are found when we require at least 3 and 5 similar P-waves trains, respectively. The occurrence of correlated earthquakes is widespread across northern California with some of the highest concentrations being observed along the transform faults within the SAF system, in the LVC region, at the GGF, and at the MTJ.

Typically, the cross-correlation coefficient of seismograms measured at common stations decays for events with increasing hypocentral separation, reflecting the combined effect of variation in the type faulting and differences in the velocity structure encountered by the two rays between their respective sources and a common receiver. We characterize and quantify the decay of seismogram similarity across the various tectonic regions in Northern California by investigating the precisely located correlated events and their associated 1.5 billion P- and 0.6 billion S-wave cross correlation. Figure 5a shows P-wave cross-correlation coefficients, binned and averaged within CC intervals of 0.01, as a function of hypocenter separation for the three tectonic regions SAF, MTJ, and LVC and the area of induced earthquakes at GGF. Results for the three tectonic regions show a remarkably similar bi-exponential decay of the CC, with a steep drop-off from CC=1 to CC~0.9 between 0 and 0.3 km, and almost linear decrease to CC~0.76 at ~2.9 km, and a steep drop-off to CC=0.7 at ~3.4 km separation distance. In contrast, the CCs for the induced earthquakes at GGF show an exponential decay from CC=1 for co-located events to CC=0.7 for events separated by 2.5 km. A histogram of the number of P-wave correlations as a function of cross-correlation coefficients shows that most seismograms, regardless of the region from which they originate, correlate with coefficients between CC=0.75 and CC=0.95, with a peak at about CC=0.79 (Figure 5b). S-wave coefficients break down faster with increasing hypocenter separation compared to the P-wave coefficients. But, similar to P-waves, S-wave coefficients for events at GGF break down fastest with increasing hypocentral separation.

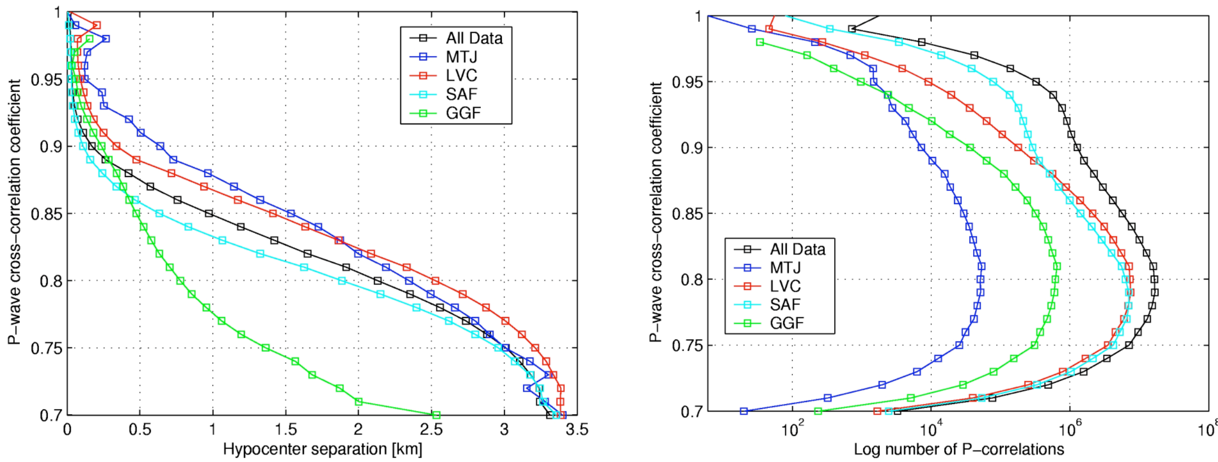

Figure 5 1.5 billion P-wave cross-correlation coefficients (*CC*) for 23,000,000 pairs of correlated earthquakes shown as a function of (a) hypocentral separation (after relocation) and (b) logarithmic number of correlations. The data is shown as the median CC within bins of 0.01. *CC*s are shown for all correlated earthquakes in northern CA (black curves) and for correlated earthquakes in the four focus regions (see Figure 1 for abbreviations).

An initial search revealed 24,438 repeating events (i.e., high CC, co-location, similar magnitudes) distributed across Northern California. Since the relative locations of repeating events within each sequence is shown to be virtually zero, we can assess the precision of the network locations by computing their location relative to the mean in each group. We find that 95% of the network locations are mislocated by less than 0.7 km horizontally, and less than 2 km vertically (Figure 5). Network epicenter mislocations at the 95% confidence level are largest for events in the Mendocino region (1.6 km), and smallest at the Geysers Geothermal Field (0.3 km). Maximum mislocations are 8 km horizontally and 7 km vertically. The significantly improved relative depths in the double-difference catalog are due to the additional S-wave cross-correlation times, a phase rarely picked on a routine basis at the NCEDC.

A Gaussian function fitted to the distribution of the cross-correlation delay times of the repeating events yields a standard deviation of 0.010 s for the P-waves (equal to the sampling rate of the NCSN stations), and 0.014 s for the S-waves, indicating sub-sample precision for many differential times measured from seismograms of highly correlated (i.e. co-located) events. The standard deviation of the corresponding differential times formed from the P-wave picks is 0.15 s, consistent with the expected pick uncertainty, and 1.5 orders of magnitudes less precise than the cross correlation data. Note that these metrics are derived from the original measurements before relocation. Outliers present in the long tails of the distributions are typically removed or down-weighted during the double-difference inversions (Waldhauser and Ellsworth, 2000).

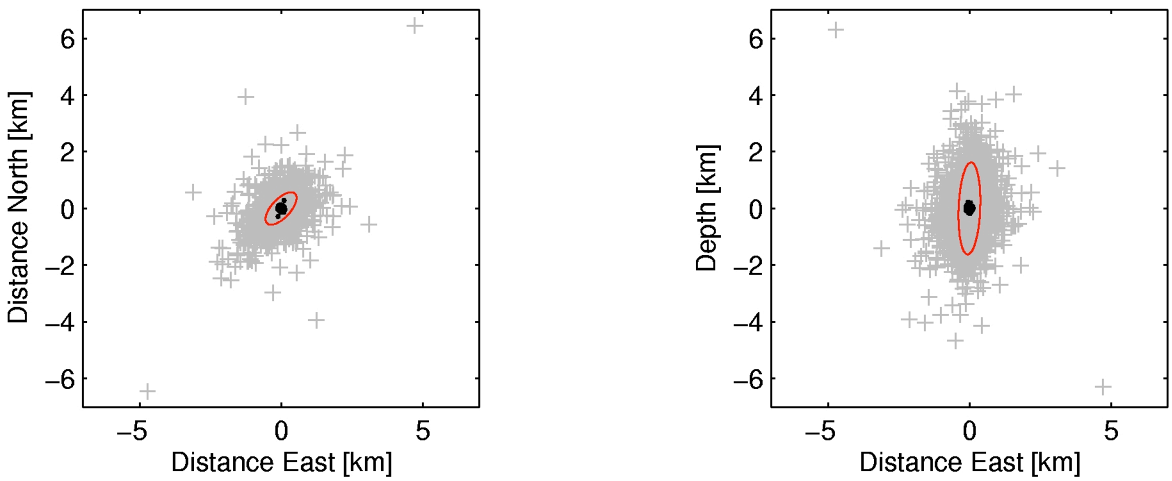

Figure 6 Map view (left) and east-west cross section (right) of a total of 24,438 earthquakes that belong to one of the 7,406 groups of repeating (co-located) events found in the relocated double-difference catalog (black +), and their corresponding locations in the NCSN catalog (gray +). Events are plotted relative to the centroid of the group to which they belong. Red solid ellipses in right panels indicate areas that contain 95% of the NCSN event locations.

4. References

Hauksson, E. and P. Shearer, Southern California Hypocenter Relocation with Waveform Cross-Correlation: Part 1. Results Using the Double-Difference Method, *BSSA,* 95, 896-903, 2005.

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

Schaff, D.P. and F. Waldhauser, Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seism. Soc. Am.*, *95*, 2446-2461, 2005.

Thurber, C., H. Zhang, F. Waldhauser, J. Hardebeck, A. Michael, and D. Eberhart-Phillips, Three-Dimensional Compressional Wavespeed Model, Earthquake Relocations, and Focal Mechanisms for the Parkfield, California, Region, *Bull. Seism. Soc. Am. 96*, 38-49, 2006.

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

Waldhauser, F., HypoDD: A computer program to compute double-difference hypocenter locations, *U.S. Geol. Surv. open-file report*, 01-113, Menlo Park, California, 2001.

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

Waldhauser, F. and D.P. Schaff, Northern California earthquake catalog relocation using cross-correlation and double-difference methods, manuscript in preparation.

5. Reports published related to this project

Waldhauser, F. and D.P. Schaff, Relocation of the Northern California earthquake catalog using cross-correlation and double-difference methods, manuscript in preparation for JGR.

Schaff, D.P. and F. Waldhauser, Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seism. Soc. Am.*, *95*, 2446-2461, 2005.

Thurber, C., H. Zhang, F. Waldhauser, J. Hardebeck, A. Michael, and D. Eberhart-Phillips, Three-Dimensional Compressional Wavespeed Model, Earthquake Relocations, and Focal Mechanisms for the Parkfield, California, Region, *Bull. Seism. Soc. Am. 96*, 38-49, 2006.

*Waldhauser, F., and D. Schaff, Massive-scale Waveform-based Double-difference Event Relocation in Northern California, *Eos Trans. AGU,* 87(36), West. Pac. Geophys. Meet. Suppl., Abstract S14B-02, 2006.

*Waldhauser,F., H. Abend, D. Bohnenstiehl, W. Kim, P. Richards, D. Schaff, M. Tolstoy, Fine-scale Seismicity of Earth's Interior: Regional- and Global-scale Double-difference Applications to Study Plate-tectonic Processes, SSA Annual Meeting, San Francisco, *Seism. Res. Lett.*, 77 (2), 2006.

Thurber, C., H. Zhang, F. Waldhauser, A. Michael, J. Hardebeck, D. Eberhart-Phillips, Three-Dimensional Compressional Wavespeed Model, Earthquake Relocations (1966- 2005), and Focal Mechanisms for the Parkfield, California, Region, *Eos Trans. AGU, 87*(52), Fall Meet. Suppl., Abstract S32A-01, 2006.

Waldhauser, F. and D.P. Schaff, Regional-scale Seismic Event Relocation in Northern California, Eos Trans. AGU, 86(52), Fall Meet Suppl., Abstract S53C-06, 2005.

Hardebeck, J. L., F. Waldhauser, A. Michael, and W. L. Ellsworth, Parkfield earthquake locations: implications for structure and fault mechanics, SSA Annual Meeting, Lake Tahoe, *Seism. Res. Lett.*, 76, 2005.

Waldhauser, F. and D. Schaff, Progress in Waveform Based Wide Area Event Relocation in Northern California, Proceedings and Abstracts, Volume XV, Southern California Earthquake Center Annual Meeting, Palm Springs, CA, September, 2005.

6. Available data and products

A beta version of the new cross-correlation based double-difference catalog (NCAeq.DDcc.1.beta) is currently being evaluated and tested (a copy of the beta catalog can be requested from the PI). We are in the process of setting up a website that will host the catalog, and tools to investigate the data online before downloading.

Contact: Dr. Felix Waldhauser
Lamont-Doherty Earth Observatory
61 Route 9W
Palisades, NY 10964
felixw@ldeo.columbia.edu