# *FY2002 Joint UCSD and Caltech Report*

# Comprehensive Application of Waveform Cross-Correlation to Southern California Seismograms

Peter M. Shearer
Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography
U.C. San Diego
La Jolla, CA 92093-0225
pshearer@ucsd.edu

Egill Hauksson and Robert Clayton
Seismological Laboratory
California Institute of Technology
Pasadena, CA 91125
hauksson@gps.caltech.edu, clay@gps.caltech.edu

## Introduction

With this project we have begun cooperation between Caltech and UCSD research into earthquake seismology in southern California. The ever-expanding waveform archive of over 400,000 local earthquake records provides an invaluable resource for seismology research that has only begun to be exploited. However, efficiently mining these data requires the development of new analysis tools, an effort that goes beyond the limited resources of individual scientists. We have begun coordinating these efforts and developing common tools and data products that can be used by us and other SCEC researchers to accomplish some of the scientific goals of SCEC.

This proposal has several purposes. These range from improving our understanding of the clustering of earthquakes and their possibly overlapping spatial and temporal source processes. We also seek to improve the earthquake hypocenters to explore the relationship between background seismicity and late Quaternary faults. Because the bulk of the background seismicity does not appear to be located on these faults, more accurate relative locations may make it possible to identify foreshocks as they happen and possibly infer the potential of a larger magnitude earthquake. As a byproduct of our analyses, we will obtain improved *P* and *S* picks to extend our data sets for 3-D Vp and Vp/Vs velocity model inversions.

At the outset of this effort, we have chosen to focus on computing waveform cross-correlation functions, because of the importance of this technique and the fact that much of the initial software development is complete. Waveform cross-correlation is an increasingly important tool for characterizing event similarity, improving earthquake location accuracy, and studying source properties (e.g., *Poupinet*, 1983; *Got*, 1994; *Dodge et al.*, 1995). Below we present some of our exciting new preliminary results.

As we head into this project several practical issues are emerging regarding data storage and processing. So far we have chosen to work with three large data sets of waveform data (Figure 1). This has given us experience in handling our data and identifying the necessary software modifications. From this experience we plan to define a modified processing thread as discussed in the proposal.

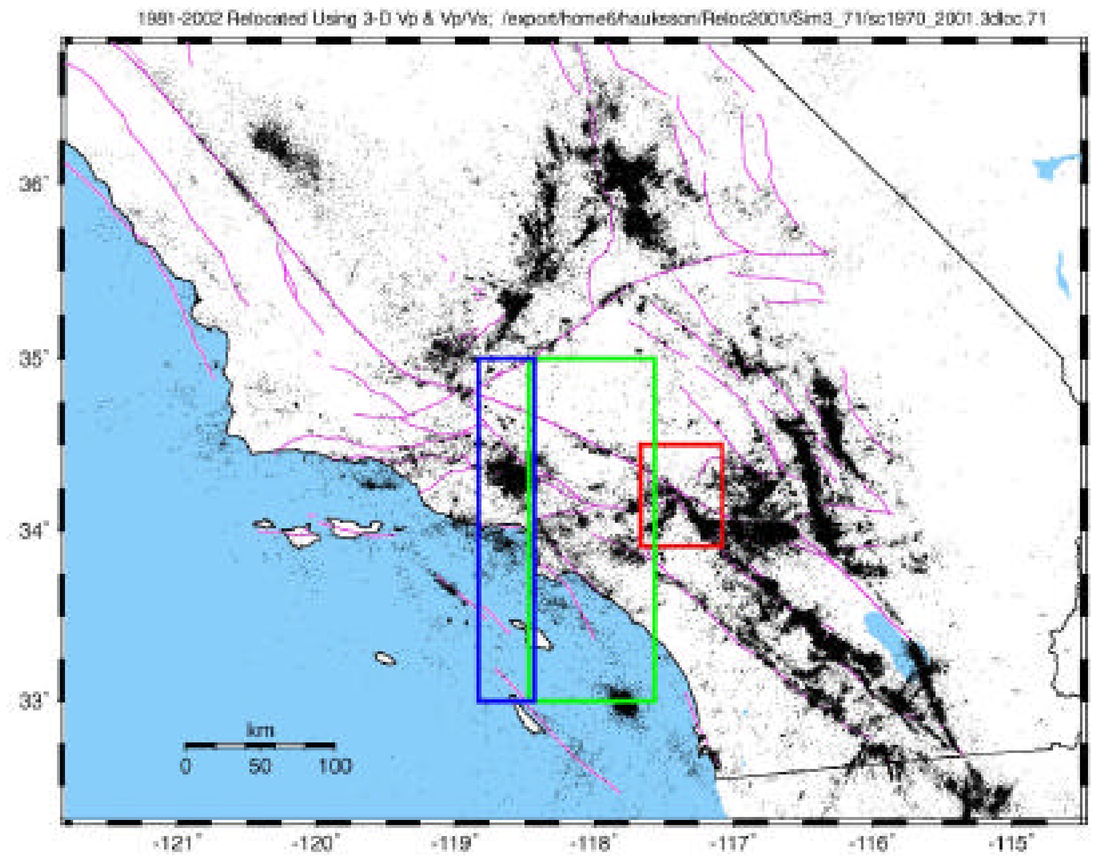


Figure 1 Map of southern California seismicity showing the study areas where data are being analyzed.

## Data and Data Processing

We have begun the joint Caltech and UCSD analyses of the vast waveform archives of the Southern California Seismic Network (SCSN/TriNet). Our initial goal is to develop and implement a system for computing waveform cross-correlation on existing and future southern California seismograms. We plan to use the results of the cross correlations for our research projects, as well as to make the results readily available to the community.

We have used the small (70 event) 1992 Yorba Linda sequence as a test for the processing steps and the algorithms that are being applied. Sample seismograms and cross-correlations from the 2002 Yorba Linda sequence are shown in Figure 2. Seismograms from two events recorded at several stations are shown along with the respective cross-correlations. The red horizontal line represents a coherency of 0.6. The M4.8 event was the mainshock and the M2.8 event was an immediate aftershock. Although both are close in space and time, their waveforms are somewhat dissimilar but can nonetheless be cross-correlated well enough to obtain differential time measurements.

Initially our intent was to first convert all the 1.4 billion seismograms from the 400,000 earthquakes in the SCSN/TriNet database into the appropriate format, and then to perform the cross-correlations. Following some tests, we decided that this was not practical because the RAID we had purchased only could store 0.8 Tbytes, and we needed about 1.4 Tbytes of disk space to hold the complete data set. Also, we were still in the process of deciding how to window and process the data before the cross-correlation.

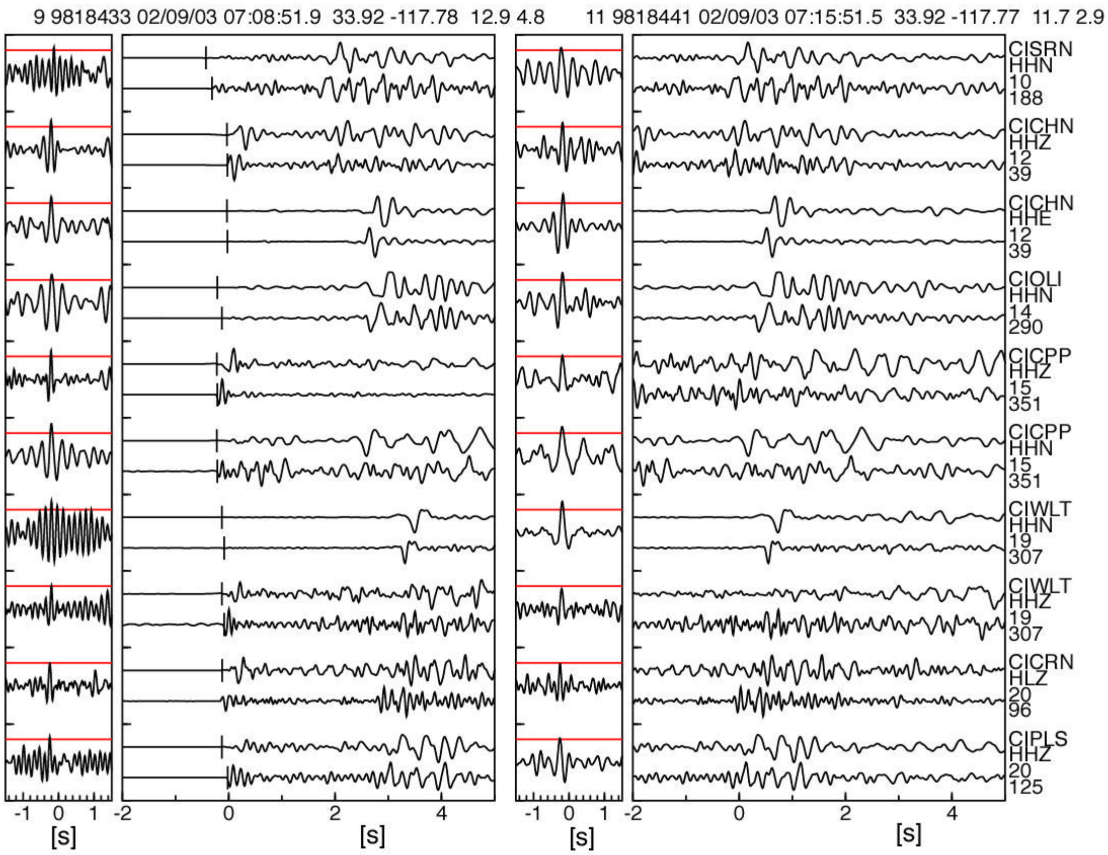


Figure 2 . Seismograms and cross-correlations from the Yorba Linda events, red line represents coherency of 0.6. Vertical bars represent phase picks.

To proceed with data processing we decided to select three regions to begin the cross-correlation project; 1) the Cajon Pass area, where the San Andreas and the San Jacinto faults merge; 2) the Los Angeles basin; and 3) the Ventura basin, including the Mw6.7 1994 Northridge earthquake. We have varied the width of each region so that it contains about 10,000 to 20,000 events. Below we list the major data processing steps:

1) Reformat and filter the waveform data
    a. Time window cut-off
    b. Spatial cut-off
    c. Re-sampling to a constant 100 Hz sampling rate
2) Determine the nearest-neighbor population of 100 events
3) Determine the cross-correlation coherency and time differential
4) Relocate using the double difference method (HypoDD), where both phase picks and cross-correlation picks included
5) Relocate using clustering algorithms of Shearer (1997)

These steps have been applied to the 2002 Mw4.8 Yorba Linda sequence and the Cajon Pass --- San Bernardino data, as discussed below.

## Results: Yorba Linda

The $M_L$4.8 mainshock occurred at 00:08 am on 3 September northeast of Yorba Linda in Orange County at a depth of 9 km. It was preceded by two foreshocks at 09:50pm ($M_L$2.6) and 10:23pm ($M_L$1.5) on Sept 2$^{nd}$. It was also followed by 23 aftershocks during the next 9 hours, with the two largest aftershock of $M_L$2.8 at 00:15am and 04:28am. An additional 50 aftershocks occurred from 4 Sep. to 30 October 2002.

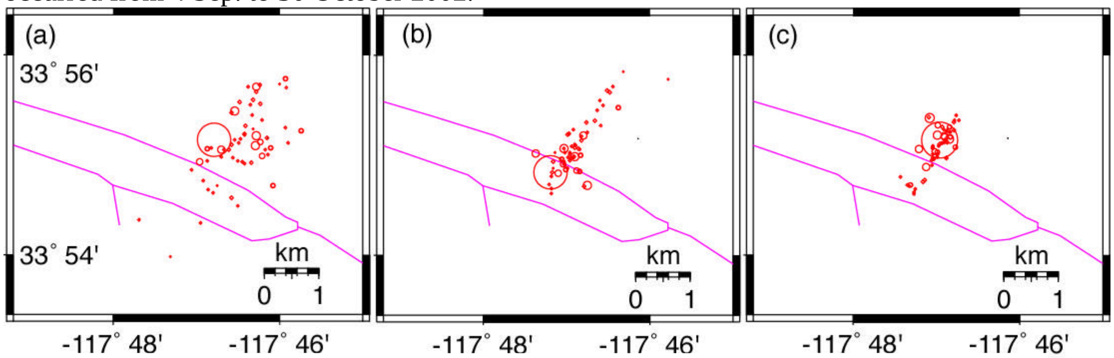


Figure 3. The ML4.8 September 2002, Yorba Linda sequence; (a) catalog locations, (b) 3-D Vp and Vp/Vs models used to refine locations, and (c) cross-correlation travel times used in the double difference method.

The mainshock focal mechanism is consistent with either a N30˚E or a N60˚W trending vertical strike-slip fault. The latter is nearly parallel to the local strike of the Whittier fault (N70˚W)one of the fastest moving faults (~2 to 3 mm/yr) in the Los Angeles basin. However, our relocations derived from waveform cross-correlation for the aftershocks appear to form a northeast trend, thus suggesting that this sequence is occurring on a small conjugate fault, adjacent to the Whittier fault or involves a geometrical complexity, such as a jog in the Whittier fault itself. We have used the Yorba Linda set of 70 earthquakes for testing and proof of concept.

The prominent clustering of the Yorba Linda aftershocks shows that through applying these methods the hypocenters are converging to a tighter cluster at each iteration step. The aftershocks clearly define a northeast trend with the mainshock located toward the center of the distribution. Previous locations suggested that the mainshock was at the end of this zone.

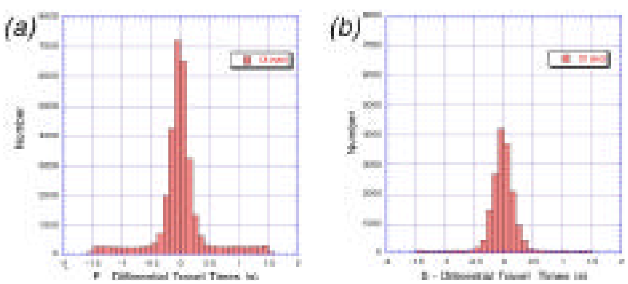


Figure 4. Histogram of differential travel times for the 2002 M4.8 Yorba Linda sequence.

A histogram of the P and S travel time differentials is shown in Figure 4. This histogram illustrates that P-waves in general correlate better than the S waves and generate more differential travel times. The half width of each distribution is similar suggesting that the differential times are of similar size.

## Results: Cajon Pass

We have chosen the Cajon Pass – San Bernardino region where the San Jacinto, San Andreas faults, and the Cucamonga faults merge as a test region for our project. This region has a high rate of seismicity that always has appeared diffuse in maps of the standard catalog locations (Figure 5). In particular, the north end of the San Jacinto fault and the Fontana seismicity trend are known for their high level of spatially scattered seismicity.

We have relocated events in this region using two different techniques. First, we relocated the seismicity using the techniques of Shearer (1997) (see figure in new proposal). This technique allows us to identify individual clusters within the background seismicity. Numerous spatially tight clusters can be seen within the clusters of background seismicity. Second, we relocated the seismicity using a refined 5 km grid Vp and Vp/Vs velocity model (e.g. Hauksson, 2000). Further, we used the new hypocenters and associated phase picks, and the cross-correlation differential travel times as input into the double-difference algorithm (HypoDD, Waldhauser and Ellsworth, 2000). Approximately 80% of the events cluster in this algorithm.

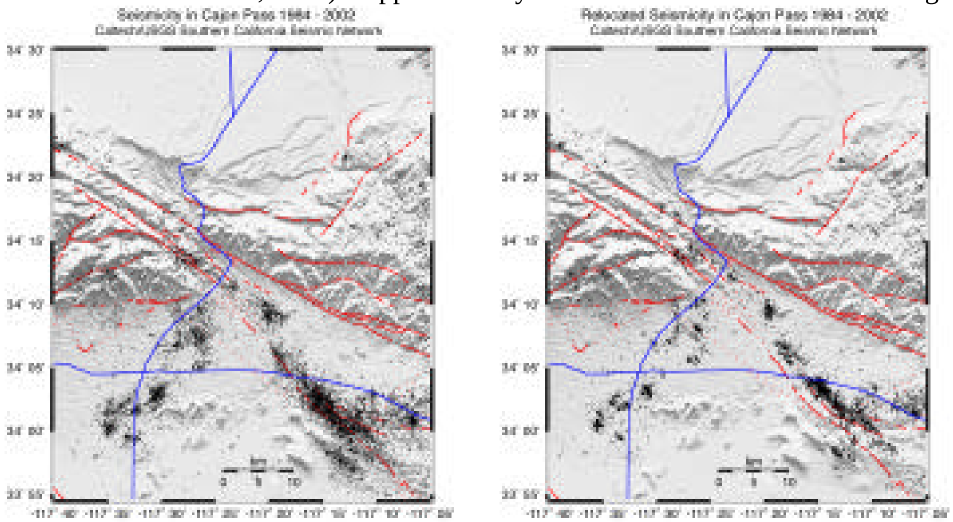


Figure 6. (a) The SCSN catalog and (b) the relocated seismicity in Cajon Pass – San Bernardino. Blue lines are freeways and red lines are active late Quaternary faults.

The final hypocenters show a much tighter clustering along the east side of the San Jacinto fault and along the Fontana trend (Figure 6). In addition, the major strands of the San Andreas and San Jacinto faults appear to be mostly devoid of background seismicity. Some of the biggest improvements in the clustering occur along the Fontana trend, striking to the southwest across Chino Valley. The relocated hypocenters in the Fontana seismicity trend form many short northwest striking seismicity alignments, sub-parallel to the strike of the San Jacinto fault to the east. These alignments could not be identified before relocation.

The major scientific issues that we plan to address with these relocated events are the merging of the San Andreas, San Jacinto, and Cucamonga faults and how the associated tectonics form major geographical features such as the Eastern San Gabriel Mountains and the San Bernardino and Chino basins. Understanding the tectonics and seismicity of this region is important, because it is one of the fastest growing areas of urban development in California. Also, the Cajon Pass is one of two major utility corridors that connect southern California, including the Los Angeles basin, to the outside world.

## References

Dodge, D.A., G.C. Beroza and W. L. Ellsworth, Detailed observations of California foreshock sequences: Implications for the earthquake initiation process. *J. Geophys. Res*., **101**, 22,371-22,392, 1996.

Got, J.-L., J. Frechet and F.W. Klein, Deep fault geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. Geophys. Res*., **99**, 15,375-15,386, 1994.

Hauksson, E. , Crustal structure and seismicity distributions adjacent to the Pacific and north America plate boundary in southern California, *J. Geophys. Res* , **105**, 13,875-13,903, 2000.

Poupinet, G., W.L. Ellsworth and J. Frechet, Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California, *J. Geophys. Res*., **89**, 5719-5731, 1984.

Rubin, A.M., D. Gillard and J.-L. Got, Streaks of microearthquakes along creeping faults, *Nature*, **400**, 635-641, 1999.

Shearer, P.M., 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, 1997.

Waldhauser, F. and W.L. Ellsworth, A double-difference earthquake location algorithm; method and application to the northern Hayward Fault, California, *Bull. Seismol. Soc. Am*., **90**, 1353-1368, 2000.

Waldhauser, F. and W.L. Ellsworth, Fault structure and mechanics of the Hayward fault, California, from double-difference earthquake locations, *J. Geophys Res.*, **107,** 10,1029, 2002.