Final Technical Report

USGS/NEHRP Award No: G09AP00067

Analysis of Microseismicity at Parkfield, California, through Improved Detection and Location

Principal Investigator: David Schaff
Co-PI: Felix Waldhauser

Lamont-Doherty Earth Observatory of Columbia University
Palisades, NY 10964
Tel: (845) 365-8826; Fax: (845) 365-8150
dschaff@ldeo.columbia.edu

Project Period: 06/01/2009 – 09/30/2010

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G09AP00067. 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

5076 earthquakes (*M* 0.2 to 6) from a local high resolution base catalog at Parkfield, California, are used to study the comparative performance of a correlation detector and standard energy detector on a sparse regional network of continuously operating stations. 86% of the events detected by a standard energy detector can also be detected by cross correlation. Correlation detection is able to find additional events by lowering the detection threshold by about 1 unit beyond what standard processing detects for Parkfield, a factor of 10 increase in number of events like Gutenberg-Richter predicts. Synthetic tests on 78,028 focal mechanisms indicate that statistically significant detections are still triggered for strike variations as large as 55°, dip variations as large as 70°, and rake variations as large as 80°. This allows for much greater variability in mechanism for co-located events than was previously thought. Application of the correlation detector was then performed on the continuous data streams of the borehole network to look for new events. One day records were requested for one month before and after the 2004 Parkfield earthquake. 10 master events were chosen from the base catalog which detected the largest percentage of the existing catalog. To estimate false alarm rates we ran the detection codes with time-reversed versions of the master templates as well. Similarity thresholds for correlation detection triggers were set for a 50% false alarm rate at each station. Then a correlation associator was created based on the expected arrival times from the master event. Adding the times of the correlation peaks reduced the false alarms to a few percent. An estimated false alarm rate of 1.13% was chosen allowing 6827 events to be detected, associated, and located as part of the new catalog. The existing catalog had 1136 events for the same time period (over a factor of six times as many). Additional events were found in the early aftershock period. The aftershocks are observed to follow Omori's law. Preliminary results seem to indicate no new foreshocks were detected.

1. Overview

This report covers the activities performed between June 1, 2009 (start date of the project) and September 30, 2010 on the application of correlation detectors for improving and analyzing the catalog of microseismicity at Parkfield, California. The work described here has been undertaken by the principle investigator David Schaff and by co-PI Felix Waldhauser. Sections 2.1 and 2.2 describe ongoing research in the field of correlation detectors performed under the auspices of both Air Force Research Laboratory contract no. FA8718-05-C-0022 and U.S. Geological Survey award no. G09AP00067. Research was initiated for the AFRL project and continued for the USGS project culminating in combined results for two publications (Schaff, 2010; Schaff and Waldhauser, 2010). Since these are already published papers this report only presents the main result with one graphic along with a short description. Section 2.3, however, goes into substantially more detail for this report since the results are in preparation for publication.

2. Investigations undertaken

2.1 One magnitude unit reduction in detection threshold at Parkfield.

Results were preliminary for Parkfield at the time of writing the USGS proposal for this project. Completion of this work was supported in part by USGS award no. G09AP00067. Results were finalized and made ready for a publication quality paper in a peer reviewed journal (Schaff and Waldhauser, 2010). For this study we are interested in quantifying the reduction in detection threshold achievable by cross correlation compared to standard procedures to extend the level of completeness of the catalog to lower magnitudes. We start with the local high resolution catalog of Thurber et al. (2006) which contains 8993 events with magnitudes ranging from 0.2 to 6 (subsequently referred to as the PKF catalog). The events were initially identified by the Northern California Seismic Network (NCSN) of short period stations. Waveforms were then requested from the IRIS DMC for seven nearby regional continous broad band stations. 5076 events out of the 8993 had waveforms available.

We compute all possible correlations using every event as a master event against all the other events within reasonable distance thresholds to quantify the best results obtainable from a complete library of master templates. P-, S-, and Lg-waves were all analyzed with 15 s, 20 s, and 50 s windows respectively and 10 s, 15 s, and 30 s lags searched over respectively forward and backwards for the time-domain cross correlation. Lg-waves are high-frequency, multiply reflected, large amplitude arrivals on regional seismograms with long durations coming in after the S-waves. The windows were centered based on the predicted travel times for the three waves. The windows are chosen to be long to increase the significance of the cross correlation values with a higher time-bandwidth product. The filter bands were 0.75 to 2 Hz, 0.5 to 3 Hz, and 0.5 to 5 Hz respectively. All pairs of events that occurred within 6 km of each other were considered amounting to 53 million correlations and about 2 weeks of continuous processing time on a four processor computer.

We employ the same method for detection purposes of using a “scaled cross correlation coefficient” (SCC) on cross correlation traces that are averaged across all three components for the particular phase to enhance the detection signal by constructive interference that was utilized in Schaff (2008, 2009) to determine a trigger whether it be a true detection or a false alarm.

Each point in the cross correlation trace, CC_i, is scaled by the mean absolute value of the moving window (length N) n points before the point

$$SCC_i = \frac{CC_i}{\frac{1}{N}\sum_{j=i-N-n}^{i-1-n}\left|CC_j\right|}$$

We use a moving window of 20 s and choose n to be 4 samples to avoid side-lobes of the cross correlation function. Intuitively, SCC is a measure of the statistical significance of a detection spike since it quantifies the deviation of the cross correlation coefficient from an empirical distribution of background values based on a moving window throughout the correlation trace.

We estimate false alarm rates with an empirical approach by applying the algorithms in an identical manner with the same parameters except for shifting the windows 120 s before the expected P-wave arrival. The idea behind this is that the windows should contain only noise and so any trigger for a given threshold would then be considered as a false alarm. This is the most robust method we know of for estimating false alarms for our comparisons since all the stations and event pairs are the same and the windows are centered on the noise characteristic right before the signal comes in for each individual station.

We decide based on an initial screening of all seven stations and phases to use only Lg-waves at stations SCZ and ISA. We further require a selection criteria of event separations less than 1 km, lags searched over to be 0.3 s, and SCC >=5. We do this to maximize the amount of detections while trying to minimize the number of false alarms. Using these selection criteria we find that 1357 events out of the 5076 or 27% are detected. The false alarms using these same criteria are 1.4%, so we estimate that the true detections are approximately 25.6%.

We compare with the same simulated prototype International Data Center ("pIDC") procedures for a standard STA/LTA detector as used for China (Schaff, 2009) except counting a trigger if it occurs within 5 sec of the first arriving P-wave this time since the locations are more accurate. The "pIDC" procedures find 140 events out of the 5076. Figure 1 shows the magnitude distributions. It can readily be seen that the correlation detector finds approximately 10 times the number of events that the "pIDC" detects, which is what would be predicted from a magnitude unit reduction in detection threshold based on a Gutenberg-Richter magnitude frequency distribution. The correlation detector also finds 120 out of the 140 events that the "pIDC" detector finds or 86%. These numbers give an indication that correlation detectors may be able to detect a high percentage of the seismicity that a standard energy detector finds and also detect a substantial number of new events that were previously undetected by routine methods.

We introduce a "normalize cumulative density function (CDF)" to quantify the reduction in detection threshold from the magnitude distributions in Figure 1. To compute the normalized CDF first the detected curve (red or blue) is divided by the green curve for the base catalog. This accounts for the greater abundance of smaller events due to a Gutenberg-Richter distribution and doesn't bias the results. Then the CDF is computed as the cumulative sum of that curve. A 95% confidence lower limit on the normalized CDF for the "pIDC" detector corresponds to magnitude 2.2. For the correlation detector the 95% confidence lower limit corresponds to a magnitude 1.3 and so the reduction in threshold is 0.9 units using the normalized CDF as a measure. We find this to be a robust and intuitive way to measure reduction in magnitude

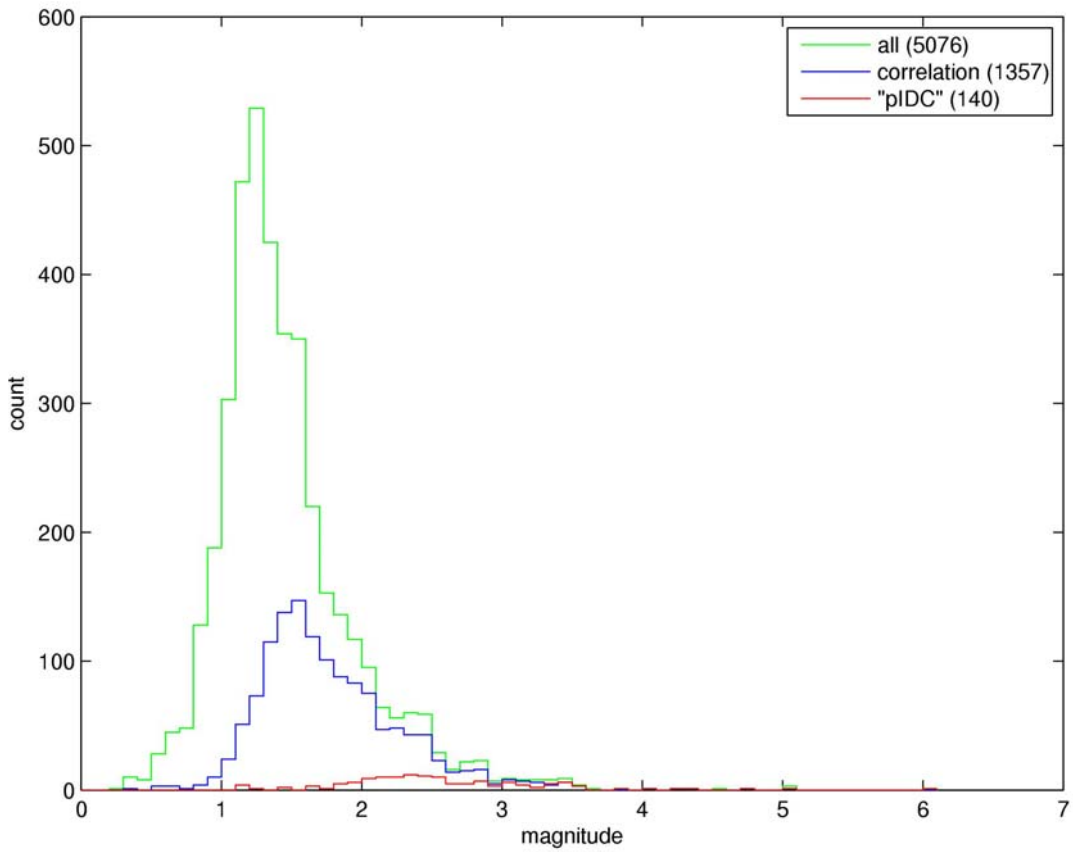

Figure 1. Magnitude distribution for Parkfield catalog and correlation and "pIDC" detectors. Number in parentheses shows total events.

threshold. Therefore correlation detectors can have approximately an order of magnitude improvement compared to standard energy detectors where there are similar events. We have demonstrated this on a large scale for Parkfield. This is compared to location where correlation combined with double difference relocations can improve locations by up to three orders of magnitude (Schaff et al., 2002).

2.2 Synthetic sensitivity tests for allowed focal mechanism variability

Schaff (2010) was part of a case study to apply correlation detectors for China initially. The paper, however, came back with major revisions requesting that synthetic tests be performed to see how variability in focal mechanisms affects the correlation results. At the time the AFRL contract had ended and the PI was being supported by the USGS grant for detection work. Since this is a fundamental question for how broadly useful correlation measurements are for real seismicity that doesn't always have identical mechisms, we thought it worthy of exploration and so added this to the initial plan of research.

Synthetic Lg-waveforms were computed at 720 km station distance with the code developed by Saikia (1994). We use a 10 layer model for P-wave velocity from 5.32 to 8.33

km/s, S-wave velocity from 3.07 to 4.81 km/s, density from 2.5 to 3.4 g/cm^3, and a quality factor for P- and S-waves of 1000 and 500, respectively. Figure 2 displays a comprehensive analysis of the average behavior of CC and SCC for many combinations of mechanisms. The top row showing dependence on rake difference is taken as the mean of 365 mechanisms for rakes in 5° increments from 0° to 360° for (strike, dip) combinations of (0°, 45°), (45°, 45°), (45°, 60°), (90°, 60°), and (120°, 30°). A very simple result emerges that the mean behavior of the vertical, tangential, radial, and average CC decays linearly from unity to -1 for rake differences of -180° and 180°.

Although individual mechanisms do not demonstrate this linear and symmetrical relationship, the average does. In fact, any fixed single (strike/dip pair) averaged over all the rakes from 0° to 360° shows this triangular function dependence on rake difference. While surprising at first, it makes sense that for rake differences of -180° and 180° the slip direction is opposite to 0° and therefore the polarity is reversed and the waveforms are exactly anti-correlated. The zero crossings are at -90° and 90° rake difference where the sense of slip is orthogonal to 0° and therefore the CC is zero on average. Also it was not expected for this decay to be linear with greater rake difference on average. Prior to this analysis we anticipated that the similarity would degrade much more rapidly with variations in focal mechanism. The SCC = 6 values (the threshold used in Schaff (2008) shown to have very low probability of false alarms of about one per day) occur at -80° and 80° rake difference which allow for much greater deviation in rake than we expected could be used for detection purposes. The average CC values corresponding to -80° and 80° are about 0.15 which are quite small so the waveforms are not nearly identical. But the SCC of 6 which measures the statistical significance of these CC values in terms of their distance from the background values indicates that the time-bandwidth product is sufficiently large to cause discernible detection spikes.

These values constructively interfere to the nearest sample on the three components for the average CC which doesn't occur for events with different paths. In these synthetic tests we are able to fix the path and only examine the dependence on focal mechanism variation. Another thing to point out that while these variations allowed for detection are larger than expected for Lg-waves in terms of how SCC behaves quantitatively, we note that for a simple wavelet for a P-wave in a homogenous halfspace, the CC value should alternate between 1 and -1 depending on what quadrant the station is in and so variations in rake from -90° to 90° should produce detections. Here we are testing a simplified 1D velocity model and the Lg-wavetrain which is composed of multiple arrivals leaving different parts of the focal sphere and that is the reason for the decay in CC as a function of rake difference.

The second row of Figure 2 performs a similar analysis of average behavior as a function of strike difference. Mean CC and SCC is shown for 27,010 mechanisms with dips ranging from 0° to 90° in 10° increments, rakes ranging from 0° to 360° in 10° increments, and strikes in 5° increments from 0° to 360°. The "W" pattern of average CC again makes sense. Consider a vertical strike-slip fault . For a strike difference of -180° and 180° the radiation pattern is identical to that for 0° and so the correlation is unity. For a strike difference of -90° and 90° the radiation pattern is opposite and so the polarities are reversed given a CC of -1. For strike differences of -135°, -45°, 45°, and 135° the radiation pattern is nodal compared to that of 0° and so the correlation is zero. The SCC threshold of 6 corresponds to strike differences of -60° and 55° allowing for significant variation in strike to produce detections although not as much as seen for rake changes. The red line in the SCC panel shows the minimum value (compared to mean) as a function of strike difference which still allows for substantial variation. The bottom

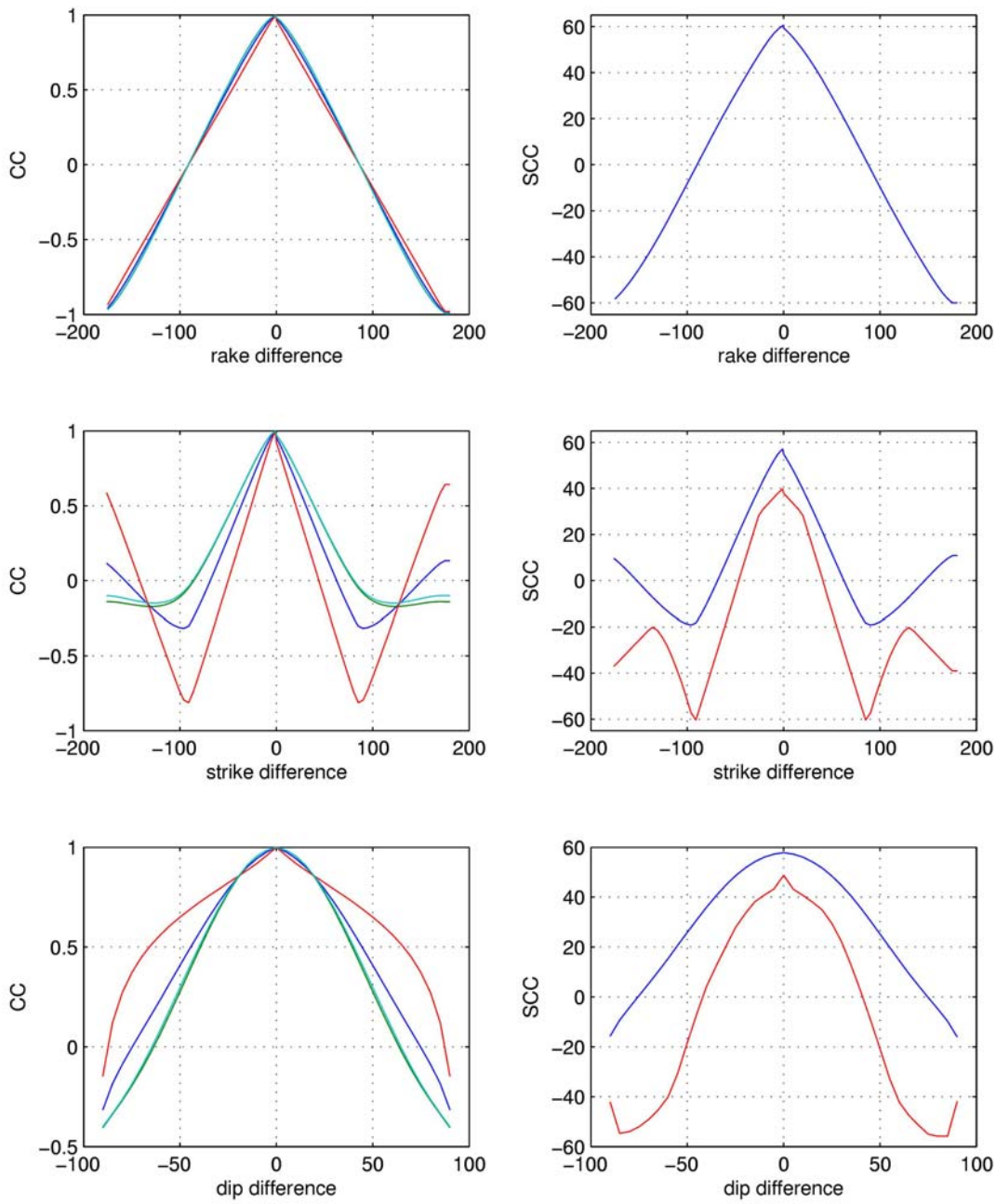

Figure 2. Mean behavior of CC (left) and SCC (right) as a function of rake, strike, and dip differences for 365 mechanisms top row, 27010 mechanisms middle row, 50653 mechanisms bottom row. CC legend is vertical (green), tangential (red), and radial (cyan) and average (blue). SCC legend is mean value (blue) and minimum value (red).

row of Figure 2 shows the mean behavior as a function of dip difference for 50,653 mechanisms with strikes ranging from 0° to 360° in 10° increments, rakes ranging from 0° to 360° in 10° increments, and dips ranging from 0° to 90° in 5° increments. Since dip angle does not continue around the complete 360° there are not equal numbers of data points entering the mean as a function of dip difference (the least are at the ends and the most at the center 0°). The SCC = 6

threshold corresponds to dip differences of -70° and 70° which is in between the strike and rake variation limits. In summary much greater variability in mechanism is allowed than previously thought for co-located events using the statistically significant measure of SCC = 6 even though the corresponding CC values can be quite low.

2.3 Application to contiunous borehole datastreams at Parkfield.

The initial work at Parkfield was based on existing catalogs of events obtained from the NCSN and windowed data. The goal was to determine what percentage of the seismicity could be detected by cross correlation and to quantify the reduction in detection threshold. This section describes the application of the correlation detector to continuous data streams to look for new events a month before and after the 2004 M 6.0 Parkfield earthquake.

2.3.1 Data Acquistion

Data is acquired from the Northern California Earthquake Data Center (NCEDC) for the Parkfield High Resolution Seismic Network (HRSN). The 13 borehole stations have network code "BP" (see http://www.ncedc.org/ncedc/station.info.html for more information). Data was requested for DP1, DP2, and DP3 channels (one vertical and two horizontal components) which are sampled at 250 Hz. Continuous one day records were requested from August 26, 2004 (julian day=239) through October 27, 2004 (julian day=301). For some stations and days the continuous records had data gaps and were not complete. This split up the data files into many fragments. No attempt has currently been made to stitch together these fragments to produce one day records where data gaps are filled with zeros. It has not been tested how filling with zeros may produce false alarms for a correlation detector. Since the fragments are quite numerous running the correlation detector on individual fragments at this stage would require a fair amount of bookkeeping. Therefore as a first pass through the data to see how the correlation detector performs we only retain records for stations that have a complete (or nearly complete) one day of continuous recording on all three channels. Table 1 shows the data coverage for the network over the observation period. Columns are julian days for which a full day's record is available at that station. If that day is not complete a zero is entered as a place holder. The Parkfield earthquake occurred on September 28, 20004 (julian day=272). From julian days 269 to 291 at least 12 of the 13 stations were operating each day. The continuous data streams were filtered from 1.5 to 120 Hz before running the correlation detector. Other than that no quality control of the continous data was performed.

2.3.2 Selection of Master Templates

Correlation detectors make use of *a priori* information in the form of master templates to be used for the matched filters. Therefore an existing archive of waveforms for previously identified events in the region of interest is needed. These events can be detected and located by conventional means and then used later by correlation applications. We again use the PKF catalog of 8993 events as our base catalog. We then made a data request from the NCEDC for windowed waveforms from the continous data streams of the HRSN borehole stations for these 8993 events. Out of these 4828 events had waveform data. This is due to the fact that the HRSN

Table 1 Continuous Data Coverage for HRSN

CCRB	EADB	FROB	GHIB	JCNB	JCSB	LCCB	MMNB	RMNB	SCYB	SMNB	VARB	VCAB
239	0	239	239	239	0	239	239	239	239	239	239	239
240	240	240	240	240	240	240	240	240	240	240	240	240
241	241	241	241	241	0	241	241	241	241	241	241	241
242	242	242	242	242	0	242	242	242	242	242	242	242
243	243	243	243	243	243	243	243	243	243	243	243	243
244	244	244	244	0	244	244	244	244	244	244	244	244
245	245	245	0	245	245	0	0	245	0	0	0	245
246	246	246	246	246	246	246	246	246	246	246	246	246
247	247	0	247	247	247	247	247	247	247	247	247	247
248	248	248	248	248	248	248	248	248	248	248	248	248
249	249	249	249	249	249	249	249	249	249	249	249	0
250	250	250	250	250	250	250	250	250	250	250	250	250
0	251	0	0	0	251	0	0	0	0	0	0	251
0	252	252	0	252	252	0	0	252	0	0	0	252
253	253	253	0	253	253	253	253	253	253	0	253	253
254	254	254	254	254	254	254	254	254	0	254	254	254
255	255	255	255	255	255	255	255	255	255	255	255	0
256	256	0	256	256	256	256	256	256	256	256	256	0
257	257	0	257	257	257	257	257	257	0	257	257	257
258	258	258	0	258	0	258	258	258	258	258	258	258
259	259	259	259	259	259	259	259	259	259	259	259	259
260	0	260	260	260	260	260	260	260	260	260	260	260
261	261	261	261	261	261	261	261	261	261	261	261	261
262	262	262	262	262	262	262	262	262	262	262	262	262
263	263	263	263	263	0	263	263	263	0	263	263	263
264	264	264	264	264	264	264	264	264	264	264	264	264
265	265	265	265	265	265	265	265	265	265	265	265	265
266	266	266	266	266	0	266	266	266	266	266	266	266
267	267	267	267	267	267	267	267	267	267	267	267	267
268	268	268	268	268	0	268	268	268	268	0	268	0
269	269	269	269	269	0	269	269	269	269	269	269	269
270	270	270	270	270	270	270	270	270	270	270	270	270
271	271	271	271	271	271	271	271	271	271	271	271	271
272	272	272	272	0	272	272	272	272	272	272	272	272
273	273	273	273	273	273	273	273	273	273	0	273	273
274	274	274	274	274	274	274	274	274	274	0	274	274
275	275	275	275	275	0	275	275	275	275	275	275	275
276	276	276	276	276	276	276	276	276	276	276	276	276
277	277	277	277	277	277	277	277	277	277	277	277	277
278	278	278	278	278	278	278	278	278	278	278	278	278
279	279	279	279	279	279	279	279	279	279	279	279	279
280	280	280	280	280	280	280	280	280	280	280	280	280
281	281	281	281	0	281	281	281	281	281	281	281	281
282	282	282	282	282	282	282	282	282	282	282	282	282
283	283	283	283	283	283	283	283	283	283	283	283	283
284	284	284	284	284	284	284	284	284	284	284	284	284
285	285	285	285	285	0	285	285	285	285	285	285	285
286	286	286	286	286	286	286	286	286	286	286	286	286
287	287	287	287	287	287	287	287	287	287	287	287	287
288	288	288	288	288	288	288	288	288	288	288	288	288
289	289	289	289	289	289	289	289	289	289	289	289	289
290	290	290	290	290	290	290	290	290	290	290	290	290
291	291	291	291	291	0	291	291	291	291	291	291	291
292	292	292	292	292	0	292	292	292	292	292	292	0
293	293	293	293	293	293	293	293	293	0	293	293	293
294	294	0	0	294	294	294	294	294	0	294	294	294
0	295	0	295	295	295	0	295	295	0	295	295	0
0	296	296	296	296	296	296	296	0	296	296	296	296
297	297	0	0	297	297	297	297	0	297	297	0	297
298	298	298	0	298	298	298	0	0	298	298	0	298
299	299	299	0	299	299	299	299	299	0	299	299	0
0	300	0	300	0	0	300	0	300	0	300	0	0
0	301	301	301	0	0	301	0	0	301	0	0	0

archive begins in 2001 and also unmet requests presumably due to data gaps.

To find the most similar events we used our correlation program designed for location purposes (Schaff et al., 2004; Schaff and Waldhauser, 2005) since it operates on segmented data and doesn't need to process as much data as the detection correlation program which needs to compute correlations for 20 s of data prior to the signal arrival to compute the SCC. Windows are centered on the theoretical P-wave arrivals based on a 1D velocity model. All possible correlations within 5 km inter-event separation distance were considered. From these measurements all values with CC ≥ 0.7 were selected as similar event pairs. The event that was similar to the most other events at this threshold was chosen as the first master event. This event and all the events that it was similar to were then removed. The second master event was then chosen in the same way as the next event that was similar to the most remaining events. Master events were then selected and ranked in terms of the highest percentage of seismicity that they recovered. Following this procedure the top ten master events correlated with 1485 other events out of the 4828 at the CC ≥ 0.7 level or 31% of the available seismicity. Including the 10 masters this is 1495 events. These events are plotted in Figure 3.

Next we test how well these 10 master events recover the 1485 events with the correlation detector code running on windowed data but computing the correlations for 20 s of data prior to the P-wave signal for the SCC statistic. We look at how the three components do individually and also the average of the three component CC and SCC traces. Table 2 breaks down these results by station. The initial thresholds that we have chosen were CC ≥ 0.7 which is typically what we use for location work (e.g. Schaff and Waldhauser, 2005; Waldhauser and Schaff, 2008) and SCC ≥ 6 (Schaff, 2008, 2010). It is seen that overall the SCC trigger detects a higher percentage of the events. Triggering on any component also detects a higher percentage of events compared to the average of three components. This is in part due to data gaps that not all events had all three components at all stations. It is also due to the fact that in general the average SCC is less than individual SCCs for a given false alarm rate. There is a fair amount of variability between stations for uniform thresholds. It is not known at this stage what the false alarm rates are for these detections. From previous work we have found taking the average SCC is a very robust way to remove false alarms because three independent measurements must align to approximately the nearest sample which has a very low probability of occurring by chance (Schaff, 2008). Therefore SCC4 is our preferred measure for detection triggers because it detects a high percentage of events while minimizing the false alarms.

Station	nev	CC4	CC123	SCC4	SCC123
BP.CCRB	1343	4%	22%	33%	64%
BP.EADB	1475	14%	37%	60%	87%
BP.FROB	1390	3%	9%	41%	75%
BP.GHIB	1452	0%	29%	42%	60%
BP.JCNB	1415	13%	53%	31%	82%
BP.JCSB	1423	3%	24%	44%	88%
BP.LCCB	1394	18%	55%	51%	85%
BP.MMNB	1392	20%	54%	44%	82%
BP.RMNB	1449	32%	82%	46%	90%
BP.SCYB	1222	19%	68%	41%	84%
BP.SMNB	1367	2%	11%	41%	62%
BP.VARB	1201	8%	37%	33%	67%
BP.VCAB	1398	13%	49%	53%	82%

Table 2. Percent of events that trigger for CC ≥ 0.7 or SCC ≥ 6. "nev" is number of events with waveforms at that station out of 1485. "CC4" is average of three component CC traces. "CC123" is any of the three component CC traces can trigger. "SCC4" is average of three component SCC traces. "SCC123" is any of the three component SCC traces can trigger.

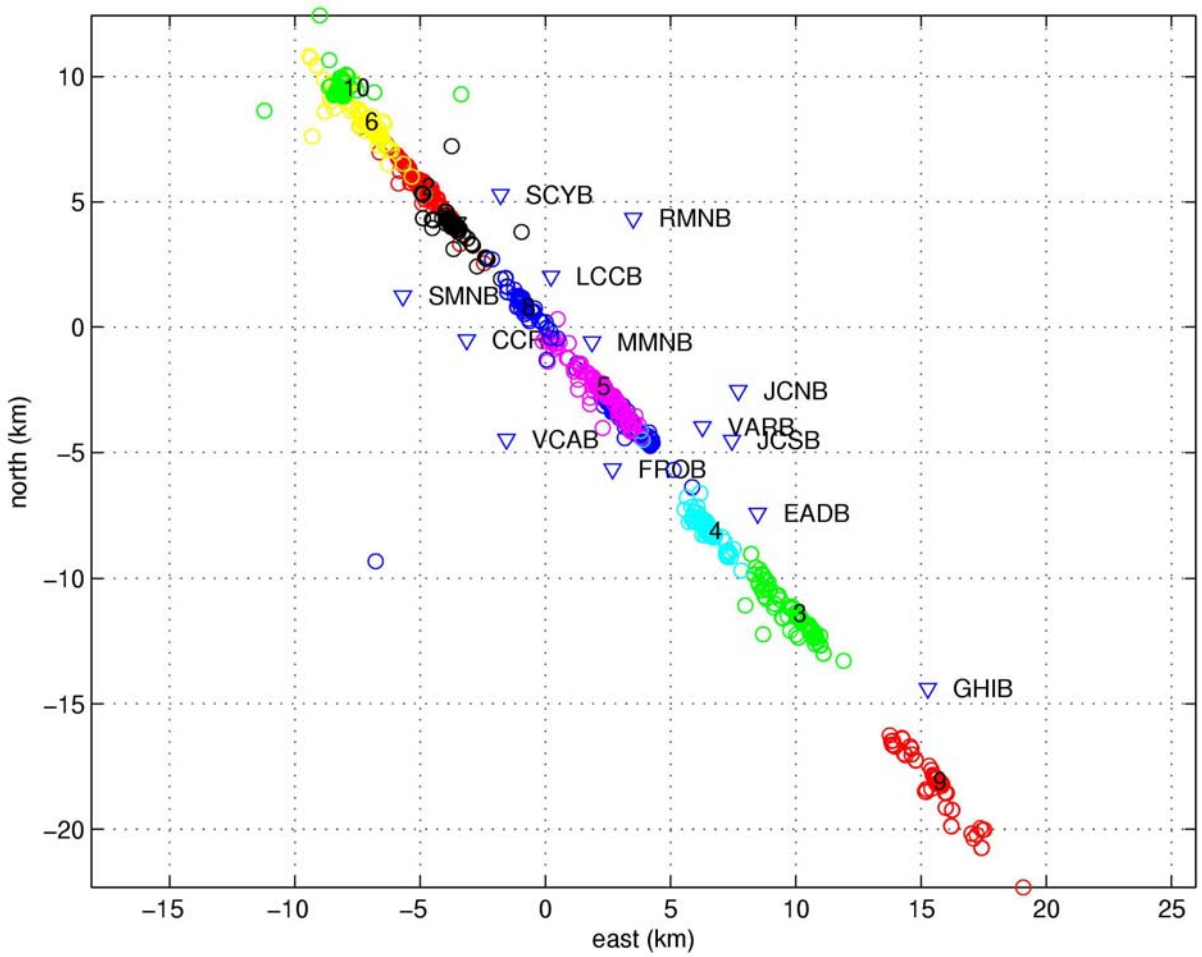

Figure 3. Locations of 13 HRSN borehole stations at Parkfield. Locations of ten master events given by text numbers 1-10. The events that correlate with each master for CC ≥ 0.7 and within 5 km inter-event separation distance are color coded.

Table 2 shows that even a single station can detect a high percentage of events (~40% for SCC4). Table 3 shows the same results if stations EADB and JCSB are combined. These two stations have waveforms available for all 1485 events at one or both stations. Now it is seen that 76% of the events are detected for SCC4 and 97% for SCC123. These results and those in Table 2 are obtained by filtering from 1.5 to 120 Hz. The second row in Table 3 shows the results for filtering from 1.5 to 15 Hz. Reducing the higher frequencies makes the CC123 exceed 0.7 on at least one component for 86% of the 1485 events. Much higher than 40% for 1.5 to 120 Hz. But the more narrow pass band causes the time-bandwidth product to be smaller reducing the statistical significance of the CC value and thus reflected in the lower percentage of events detected by SCC for a set threshold (78% for SCC123). From this we conclude that 1.5 to 120 Hz gives more statistically significant results as well as a higher percentage of detected events so these are the filter bands that we use for subsequent processing. As a point of interest row three of Table 3 examines how well these 10 master events detect all 4818 events that are available (4828 with waveform minus the 10 masters). Stations EADB and JCSB combined had data for

Table 3. Percent of events that trigger for CC ≥ 0.7 or SCC ≥ 6.

Station	filter	nev	CC4	CC123	SCC4	SCC123
EADB & JCSB	1.5-120 Hz	1485	15%	40%	76%	97%
EADB & JCSB	1.5-15 Hz	1485	27%	86%	63%	78%
EADB & JCSB	1.5-120 Hz	4816	7%	51%	80%	98%

4816 of the 4818 events. It is remarkable that only 10 masters can detect such a large percentage of 4816 events (80% for SCC4 and 98% for SCC123). SCC is what we use as a trigger for the correlation detector algorithm. It is seen that 51% are detected for CC123. This is greater than the 1495 (31%) events identified before for the correlation location algorithm presumably because any component can trigger (not just the vertical) and the inter-event distance cut-off of 5 km is not imposed.

2.3.3 Selection of Trigger Thresholds

Now that 10 masters have been identified that detect a significant fraction of the existing catalog, the next step is to run the correlation detector on the continuous data streams to look for new events that are not in the catalog. The master window lengths are 1 s. It takes approximately 45 s to run the correlation detector on our computer for one master template on a one day record on one channel. Multiply this by three components and our ~60 day observation period and this equals 2.25 hours of processing time for one master template. Multiply this by the 10 masters we have chosen and it takes 22.5 hours or about 1 day to complete one station. Multiply this by 13 stations and the total processing time is approximately two weeks for one CPU. One reason we chose only the top 10 master events as opposed to 100 or even 1,000 is because of computational efficiency. The processing time scales linearly with the number of master events. (For location work, we treat every event as a master within certain distance thresholds. For detection work this is infeasible.) The whole procedure is easily parallelized in that it can be split up over several processors according to master template or station independently. Our computations were performed on a four processor machine with 4 Gb of RAM.

For the continuous runs we compute the full SCC traces but store only triggers above a certain threshold. Since we don't know what the optimal thresholds are we choose SCC123 ≥ 5 and SCC4 ≥ 4.5. We employ a reset parameter of 1 s after the threshold is first exceeded before a new trigger can be recorded. Figure 4 shows how the number of triggers varies as a function of threshold in blue for SCC and CC at station BP.EADB for the ten master events combined. In this case the curves for SCC4 and SCC123 are about the same.

To determine the optimal thresholds we need to have some idea of what the false alarm rates are. It is seen that the curves rapidly increase for lower thresholds for thresholds approaching SCC ≥ 6 but this may be problematic if there are too many false alarms to deal with. To estimate the number of false alarms we use the same ten master templates at each station but simply reverse them in time and run them through the detection code on the continous data streams again. The idea behind this is that we are using a waveform with the same window length and frequency content both of which affect CC and SCC values, but the difference is that these templates should not correlate with any real earthquake signals present because the shape is different. The continuous data should then appear as noise compared to the template. In this

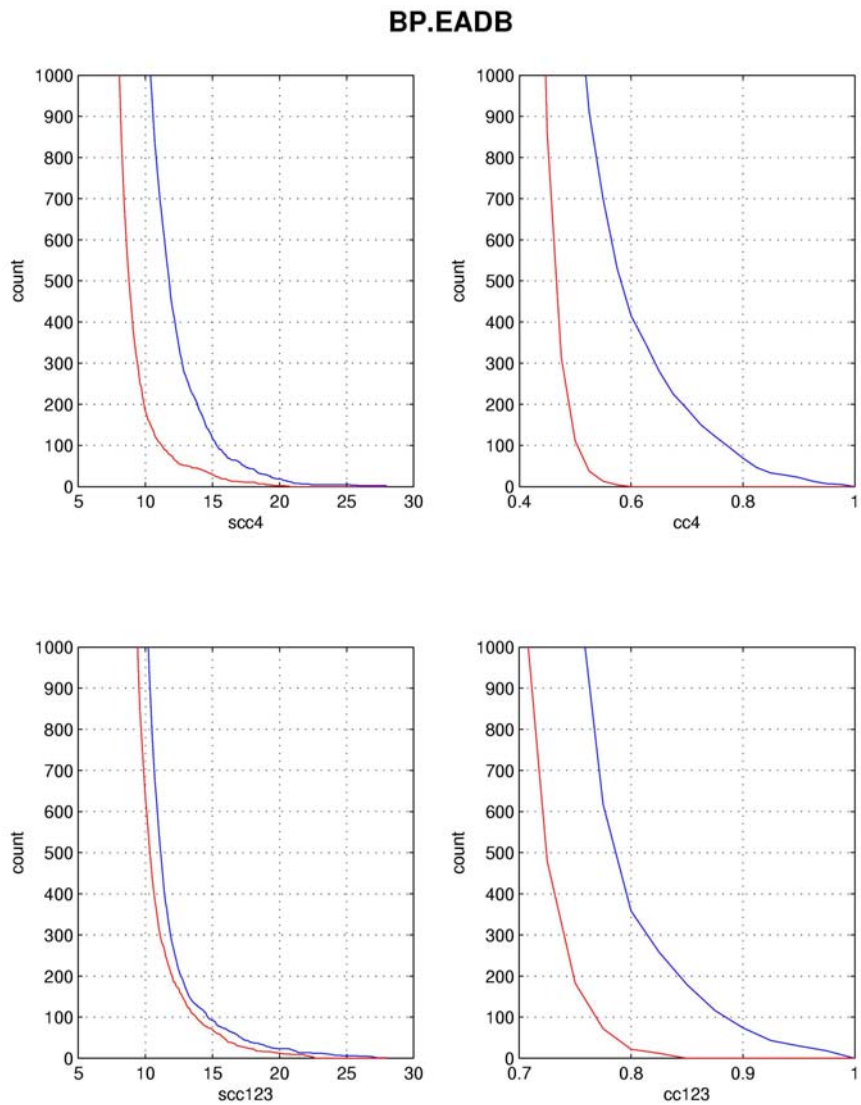

Figure 4. Number of triggers as a function of SCC4, CC4, SCC123, and CC123 for ten master templates at station BP.EADB (blue). In red are the false alarm triggers for the same ten master templates but now time reversed.

way we can empirically determine false alarms using the raw data itself and account for days that might have more noise, instrumental glitches, disimilar earthquakes, etc. Figure 4 shows how the time reversed master templates exceed certain thresholds as a fuction of SCC and CC in red. There is quite an abundance of false alarms that again grows rapidly with decreasing SCC. For SCC123 the number of false alarms is almost equal to the number of signals. Note that SCC4 eliminates many of these false alarms due to the fact that the timing of the triggers needs to align to approximately the nearest sample to constructively interfere to produce a detection spike.

To estimate the number of true detections we simply take the difference between the blue (signal) and red (time-reversed) curves. In order to estimate the percentage of false alarms we divide the red curve by the blue curve and multiply by 100%. Figure 5 plots these two values derived from Figure 4 for the different similarity measures. As expected there is a trade-off as you reduce the similarity threshold you increase the number of true detections but also the percentage of false alarms. One then needs to judge what threshold to set to allow enough true detections while keeping the number of false alarms manageable.

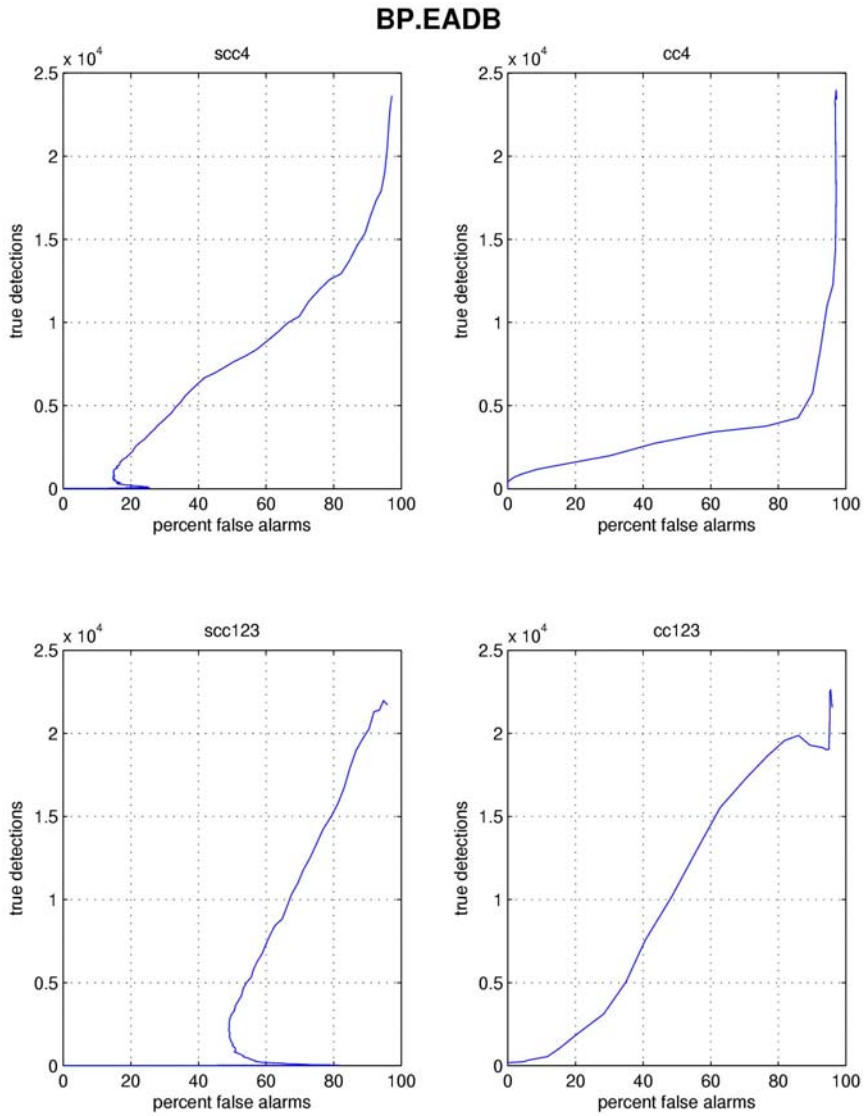

Figure 5. Trade-off of estimated number of true detections as a function of percent false alarms for SCC4, CC4, SCC123, and CC123.

We are surprised at the number of false alarms for this data set based simply on measures of similarity like SCC and CC for thresholds that have been demonstrated to have very low false alarm rates for our previous work of a few percent (Schaff, 2009; Schaff and Waldhauser, 2010). As far as we know in the literature for other correlation detector research a careful treatment of false alarms has not been performed and therefore caution must be used in interpreting the results of the detected events in case thresholds have been set too low and therefore include many false alarms. One reason for the increase in the number of false alarms is that continuous data is now being used as opposed to windowed data before which would only use the noise portion before the P-waves to estimate the false alarms. Therefore the window lengths for the waveforms that should have had signals were the same as those that should have noise. Now the window lengths for the noise (~ 2 months) are much longer than the signals. But Schaff (2008) did semi-empirical tests on 36 days of real seismic data (sampled at 20 Hz) to estimate the false alarm rates for one master and one station and came up with approximately one false alarm per day for

SCC thresholds of 6. What we have discovered, however, is that the number of false alarms for a correlation detector are approximately constant for a given number of samples, or that the false alarm rate per day scales linearly with the sampling rate. In other words, if we had approximately one false alarm per day per master for 20 Hz data then we would have about ten false alarms per day for 250 Hz data per master on these borehole stations. This is a disadvantage of correlation detectors with respect to standard energy detectors whose false alarms don't depend on sample rates. Also the number of false alarms scales linearly with the number of master templates. This is another reason why it is preferable to chose a smaller set of master events besides computational efficiency. These are some of the reasons why the number of false alarms is higher than expected. A third reason is because of disimilar earthquakes which we shall see in section 2.3.4.

Figure 6 shows the true detection versus percent false alarm curves for all the stations. Some stations are quite variable and show strange trade-offs. This makes it challenging to decide on optimal thresholds. If we are extremely conservative with the similarity thresholds (false alarm rates of ~1%) only a handful of events (~100s) get detected (most likely co-located repeating events that have regular recurrence intervals). Lowering the threshold allows events separated by greater distances from the master to be detected. Based on earlier work with correlation detectors that show one magnitude unit reduction in detection threshold is possible, we would expect several thousand to be detected (the PKF catalog has 1136 events for this time period). We decide to experiment with a uniform false alarm rate of about 50% for each of the stations shown as the red circle in Figure 6. Table 4 gives the statistics for these false alarm rates, the corresponding thresholds and estimated number of true detections. The number of true detections is quite station dependent varying from 582 at BP.SCYB to 24,352 at BP.RMNB. The value for the SCC4 threshold also shows some variability across stations. All of them except one are greater than the 6.0 threshold that we initially chose which would correspond to false alarm rates of greater than 50%. Nine of the thirteen stations, however, fall in between 5.9 to 7.0. It is not known why station BP.SCYB is so anomalous at 27.1.

station	triggers	%far	true	scc4thr
BP.CCRB	7387	51	3649	6.3
BP.EADB	15533	51	7658	6.5
BP.FROB	2175	51	1074	9.2
BP.GHIB	2013	48	1040	6.9
BP.JCNB	23726	52	11492	6.5
BP.JCSB	2203	51	1090	9.8
BP.LCCB	22258	51	11001	6.1
BP.MMNB	18923	49	9623	6.4
BP.RMNB	55898	56	24352	6.2
BP.SCYB	1158	50	582	27.1
BP.SMNB	9297	51	4589	6.3
BP.VARB	9771	51	4821	7.1
BP.VCAB	23217	48	12070	5.9

Table 4. Statistics for each station. Number of triggers for master templates above the given SCC4 threshold which corresponds to about a 50% false alarm rate (%far column) and number of estimated true detections.

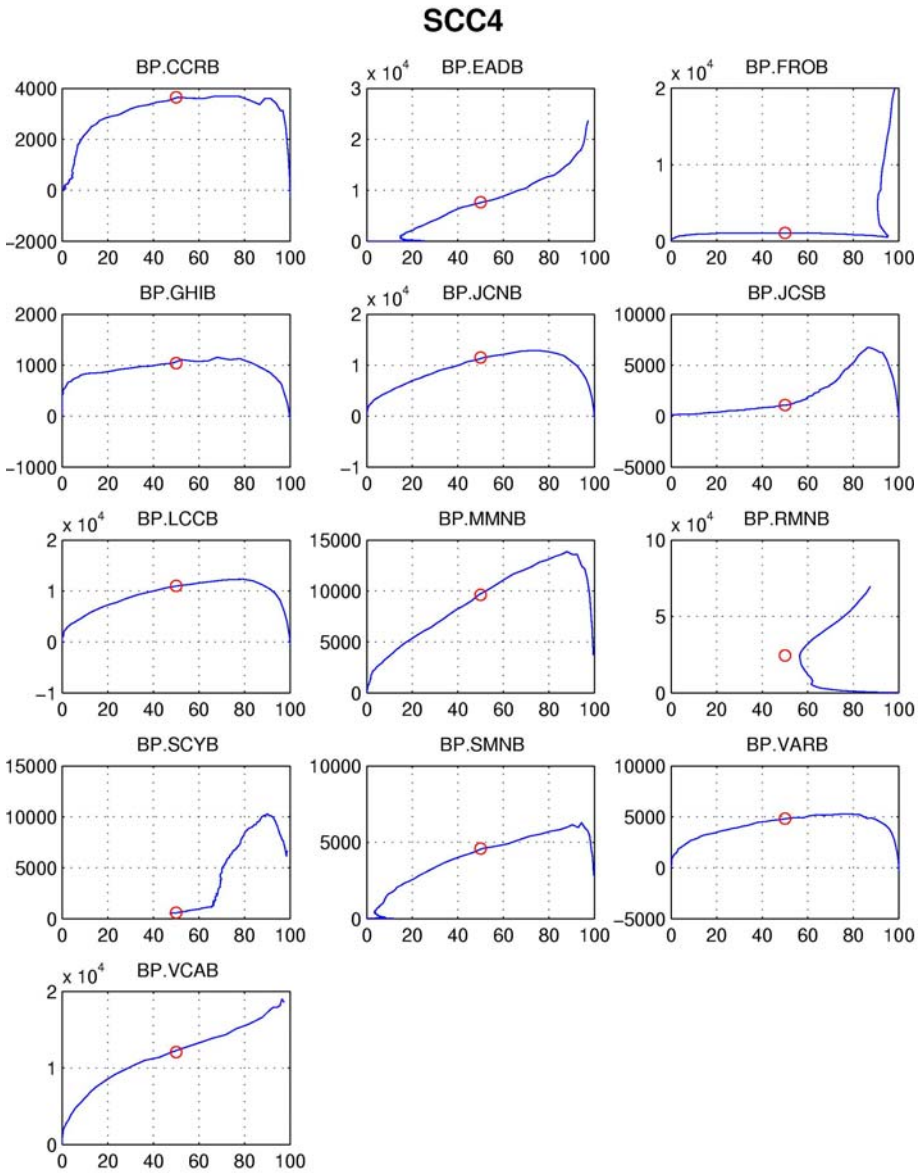

Figure 6. Trade-off of estimated number of true detections (y-axes) as a function of percent false alarms (x-axes) for SCC4 at all thirteen stations. Red circle corresponds to 50% false alarm rate.

2.3.4 Results of Correlation Detector

Using the SCC4 thresholds listed in Table 4 corresponding to 50% false alarm rates, we plot in Figure 7 the time history of the trigger times for both the master signals and time-reversed templates for each station. The x-axes are days after the Parkfield earthquake with zero corresponding to the origin time of the 2004 Parkfield earthquake, covering approximately one month before and after the mainshock. The y-axes are cumulative number of events. In blue are the origin times for the 1136 events in the base PKF catalog for the same time period.

Every station shows triggers for increased aftershock activity following the mainshock for the master signal templates. Most of the stations follow a *ln(t)* relationship corresponding to the *~1/t* decay of Omori's law. Interestingly the false alarm rates given by the time-reversed triggers in magenta are not constant over time, but also show increased activity in the aftershock

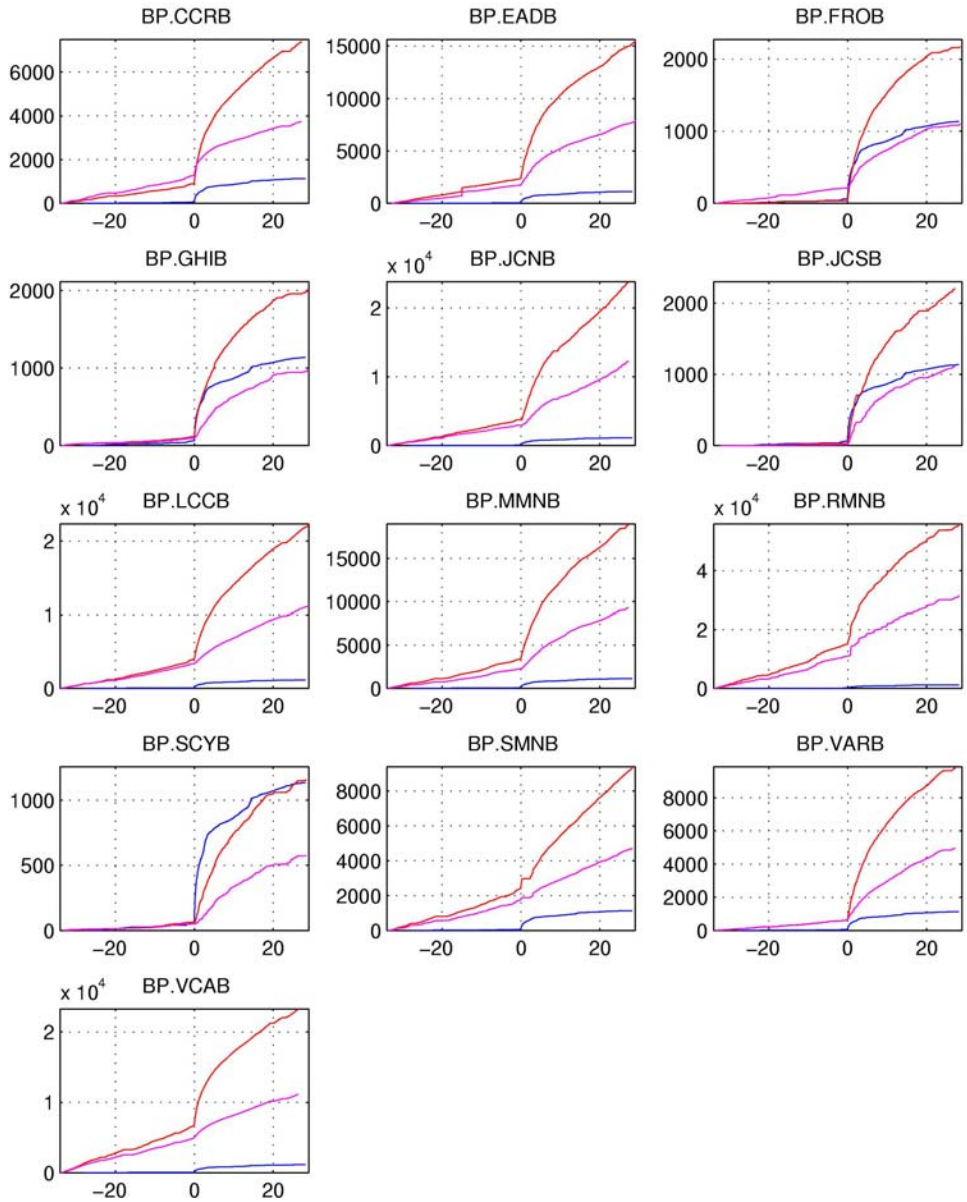

Figure 7. Cumulative number of triggers for master signals (red), time-reverse templates (magenta) and events in PKF catalog (blue). X-axes are days after the 2004 Parkfield earthquake.

period with decays following Omori's law. The reason for this is that while a correlation detector is the optimal means of detecting a known signal in Gaussian noise (Van Trees, 1968), it was shown in Schaff (2008) that relatively high CC values can occur when there are energetic but disimilar signals in the record. These is because certain high amplitude peaks and troughs of the master template and the unrelated earthquake can line up and produce high values when cross-multiplied and summed. Cross correlation values are dominated by the higher amplitudes because cross mutliplying is like squaring the amplitudes. This is a problem for earthquake signals which have variable amplitudes. It isn't a problem for Gaussian noise which has relatively constant amplitudes. Another benefit of using SCC as the measure of similarity is that it weeds out some of these false triggers on energetic but disimilar signals (Schaff, 2008). But as can be seen from Figure 7 it doesn't weed out all of them.

What this means is that our calculations of the false alarm rates are over estimates in the sense that many of the false alarms are true earthquakes and not just background seismic noise. These "detections", however, are more like those for a standard energy detector (triggers on high amplitudes) instead of a correlation detector which depends on waveform similarity. The timing of such "detections" is not precise as we will see in section 2.3.5 compared to correlation derived times of the peaks which are extremely accurate and make the association problem much easier. For these reasons we prefer to stick with our conservative estimates of the false alarm rates and true detections. In practice, for correlation detectors it becomes important to choose master events not just from any region from a large library of templates but to be selective and include the *a priori* information of location of the master event as well (not just shape of the waveform). In this way only a subset of master events is chosen for a focused region of interest.

The selection of what false alarm rate to use also depends on the rate of seismic activity. For example for long periods of quiescence there may be few true earthquakes. There are most likely few false alarms, too, because the data looks more like Gaussian noise. But the length of the time window, the sampling rate, and the number of master events all affect the number of false alarms. Therefore the percentage of false alarms, as we have computed it, may be higher for a given SCC threshold. This can be seen in the preseismic period in Figure 7 where the red and magenta lines are close to each other (which would give a higher percentage of false alarms relative to signal triggers). For BP.CCRB it is seen that the false alarms even exceed the signal triggers in the preseismic period. During periods of energetic aftershock sequences, selection of false alarm rates and corresponding SCC thresholds may need to be adjusted. In these cases there are a greater number of true events for a given time window but there are also a greater number of false alarms. Thus a uniform false alarm rate or SCC threshold may not be able to be applied to all periods with different levels of seismic activity. Further research is needed to explore how adjustments may need to be made in periods of quiescence or known intense aftershock activity. For our purposes, however, we have currently experiemented with a uniform false alarm rate of 50% for the correlation detector.

2.3.5 Results of Correlation Associator

Correlation computations not only provide measurements of waveform similarity (maximum peaks of CC) but also extremely accurate times (alignment of peaks). These times are often accurate to subsample precision (e.g. Poupinet et al., 1984; Schaff et al., 2004). The similarity values are used by a correlation detector with certain thresholds to produce detection triggers. The time values can be used by a correlation associator to identify new events. A unique feature of a correlation detector, as compared to a standard energy detector, is that it uses a master event with a known origin time and location. Since events with similar waveforms typically occur within some limited distance range from the master event an initial guess for the origin time and location of a detected event can be assigned with as few as one station triggering. In contrast, association for standard energy detectors require at least three stations with triggers.

To create a correlation associator, we compare each detection trigger at a certain station for a given master with the detection triggers close to expected arrival times for the other available stations based on the arrival times for that master. First arrival time differences between pairs of stations for the master event are calculated to remove the origin time. Then arrival time differences are computed for detection triggers at pairs of stations which removes

thes the origin time from the candidate event. Then a station arrival time delay misfit is computed as the difference between the master event station pair arrival time difference and candidate event station pair arrival time difference. If this is value is zero, the arrival times agree as expected and the candidate event is identified as an associated event that is possibly co-located with the master event. For candidate events that are separated by a certain distance from the master event this misfit will be greater than or equal to zero, depending on the orientation and length of the relative position vector of the candidate event to the master and the azimuth to the stations.

Table 5 shows the statistics for a range of misfits from 0.05 to 1 s. Greater misfit thresholds allow for events separated by greater distances from the master to be associated (misfits less than or equal to the threshold count for associations). The "stapair" column is the minimum number of station pairs required for an event to be associated. Note this is different than the number of stations that are required for an association (for example three minimum for standard energy detectors). We are using a range of station pairs instead of number of stations because it more completely captures the available information and statistics for the association stage. To run the associated events through a single event location algorithm requires a minimum number of stations, however. So the "minsta" column reflects the minimum number of stations that must have triggers for each station pair minimum. Changing the minimum station pairs to lower values allows for events of lower magnitude to be detected because they don't have to trigger at as many stations.

What immediately jumps out is that the percent false alarms, computed as the "timerev" column divided by the "signal" column multiplied by 100%, has been greatly reduced from 50% to a few percent for most of the stapair and misfit threshold combinations. And this is achieved while having a substantial number of estimated true detections. The false alarm rates increase

stapair	minsta	misfit	signal	timerev	true	%far
1	2	0.05	10474	402	10072	3.84
1	2	0.25	16560	1602	14958	9.67
1	2	0.50	20284	2824	17460	13.92
1	2	1.00	24437	4460	19977	18.25
2	3	0.05	5230	3	5227	0.06
2	3	0.25	7258	82	7176	1.13
2	3	0.50	8490	236	8254	2.78
2	3	1.00	9991	653	9338	6.54
3	3	0.05	3105	0	3105	0.00
3	3	0.25	4440	0	4440	0.00
3	3	0.50	4907	9	4898	0.18
3	3	1.00	5430	62	5368	1.14
4	4	0.05	1739	0	1739	0.00
4	4	0.25	2990	0	2990	0.00
4	4	0.50	3234	0	3234	0.00
4	4	1.00	3426	4	3422	0.12
5	4	0.05	1051	0	1051	0.00
5	4	0.25	1893	0	1893	0.00
5	4	0.50	2029	0	2029	0.00
5	4	1.00	2118	0	2118	0.00
10	5	0.05	92	0	92	0.00
10	5	0.25	116	0	116	0.00
10	5	0.50	117	0	117	0.00
10	5	1.00	120	0	120	0.00

Table 5. Statistics for correlation associator for various combinations of minimum number of station pairs, minimum number of stations, station arrival time delay misfits (see text for explanation). Number of signal and time-reverse triggers and estimated true detected and associated events and false alarm rates.

with decreasing stapair and increasing misfit. These results are encouraging because they confirm what we supposed that the timing information from the correlation measurements are extremely accurate. Consider stapair ≥ 2 and misfit ≤ 0.05 s which must have a minimum of three stations triggering. This captures an estimated 5227 number of true associated events with a false alarm rate of 0.06%. Keep in mind that average pick errors for the NCSN stations are on the order of ~0.1 s. Here we have agreement in the arrival times to less than 0.05 s and this includes any true arrival time differences of misfit due to separation of the events. Note also that only three events get "associated" for the false triggers in the time-reversed column. Therefore the high number of false alarms for the time-reverse templates using only the SCC4 similarity measure get screened out by including the extremely accurate timing information and setting a restrictive misfit parameter for the correlation associator.

We take an interest in the row for stapair ≥ 2 and misfit ≤ 0.25 s because it has a higher number of true detections (7176) while maintaining a false alarm rate of just 1.13%. Table 6 breaks these results further based on each master event. There is some variability in the false alarm rates for different master events. There is also substantial variability in the number of estimated true detections. It is remarkable that one master event can detect an estimated 3420 events within a two month time period surrounding the 2004 Parkfield earthquake, a month of which is the aftershock period.

master	signal	timerev	true	%far
1	3433	13	3420	0.38
2	270	7	263	2.59
3	1247	24	1223	1.92
4	361	16	345	4.43
5	130	7	123	5.38
6	161	2	159	1.24
7	148	2	146	1.35
8	898	4	894	0.45
9	470	3	467	0.64
10	140	4	136	2.86

Table 6. Statistics for correlation associator broken down by master event. Number of signal and time-reverse triggers and estimated true detected and associated events and false alarm rates.

2.3.6 Foreshocks and Aftershocks

We take the events identified from the correlation associator with stapair ≥ 2 and misfit ≤ 0.25 s and run them through HYPOINVERSE to get initial locations and refine origin time estimates. Of these 6827 events are locateable. The time history of the origin times is displayed in Figure 8. It is conservatively estimated that about 1% of these events are false alarms. But there still is the chance that some of these are true events as was seen before in the correlation detector section. The 1136 events in the PKF catalog for the same time period are shown in red on Figure 8 for comparison. There are over six times as many events in the new catalog. And these 6827 events were detected with only ten master events, a gain of over 600 times. 370 of the 1136 events (33%) in the PKF catalog have origin times within 10 s of the new catalog. This percentage is slightly lower than the "SCC4" column of Table 2 because we are using higher SCC4 thresholds, imposing three stations instead of one station, and restricting to 0.25 misfits. Table 2 and Figure 3 were for 5 km inter-event separation distances. A 0.25 misfit allows for a

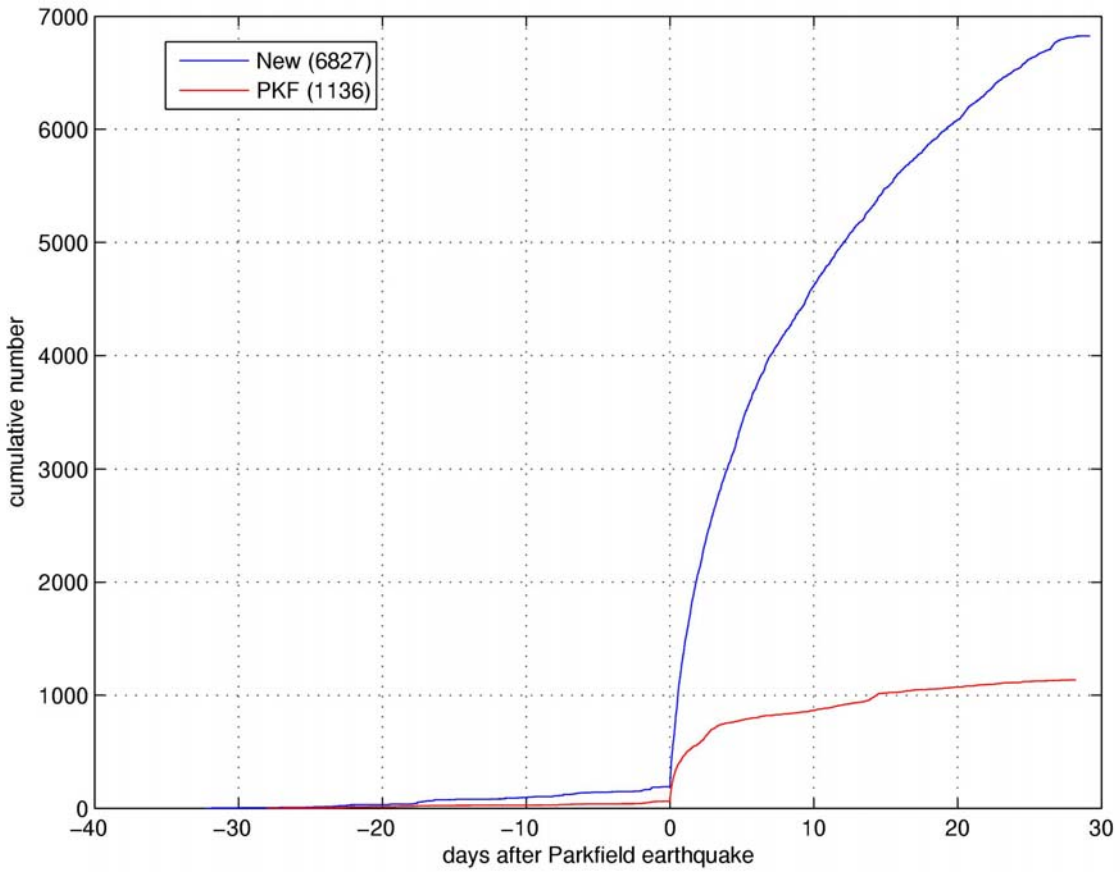

Figure 8. 6827 events detected, associated, and located in the new catalog (blue) for two months surrounding the Parkfield earthquake compared to the 1136 events in the PKF catalog (red) for the same time period.

maximum of 0.75 km separation of the associated event and master event if the two events and two stations all occur on a line. If the number of masters were increased to detect close to 100% of the PKF catalog instead of 33% then we can assume that most likely the the number of events in the new detected catalog will increase. A simple estimate would be a factor of three increase, but that is probably an upper bound because we chose the ten masters to be the ones similar to the most events in the PKF catalog.

Figure 8 displays a nice Omori's law decay for the aftershocks in the new catalog. Peng et al. (2006) discovered an anomalous early aftershock decay rate after Parkfield using a different technique of looking at high frequency envelopes. They found a deficiency in aftershock rate that was less than predicted by Omori's law. One reason for a deficiency in the completeness for routine catalogs in the early postseismic period is that so many aftershocks are occurring that the processing is overloaded and misses many events. Another reason is that aftershocks are buried in the coda of the mainshock that is still reverberating. Schaff and Waldhauser (2010) demonstrated that a correlation detector can detect aftershocks in the coda of a mainshock that are undiscernable to the eye. Recent work by other authors has also used matched filters (cross correlation) to detect early (repeating) aftershocks of the 2004 Parkfield

earthquake (Lengline and Marsan, 2009; Peng and Zhao, 2009). Table 7 shows the number of aftershocks in the first 60 minutes of our new catalog of 6827 events compared to the old catalog of 1136 events. The column labeled "First" is the time of the first aftershock after the origin time of the Parkfield earthquake. Peng et al. (2006) observed a steady rate of aftershock activity following the 2004 Parkfield earthquake in the first 130 s followed by a power law decay according to Omori's law afterwards. We observe a deficiency of aftershock activity too in the first 2 min (120 s) but since we only have two earthquakes, we can not conclude if the rate is steady.

Table 7. Number of events in early aftershock period of 2004 Parkfield earthquake

Catalog	First	2 min	5 min	10 min	30 min	60 min
New (6827)	1.6 min	2	11	22	53	109
PKF (1136)	3.7 min	0	2	8	32	64

Figure 9 zooms into the preseismic period before the 2004 Parkfield mainshock to look for any foreshock activity. The cumulative number of events plotted in the top of Figure 9 shows a gradual slope for the first 10 days. The slope steepens for the remaining 23 days. In the final three days there is a ramp up of activity approaching the mainshock. This is the steepest slope on the curve. The next steepest part occurs around day 18. The bottom of Figure 9 shows the preliminary locations of these events relative to the 2004 Parkfield M 6.0 hypocenter shown by the black star and circle at zero along strike distance. The events are color coded by time before the mainshock with dark red occurring less than five days before. The dark red events occur about 15 km away from the hypocenter of the M 6.0 event and are probably too distant to be considered foreshocks. More likely they represent an aftershock sequence of a smaller event. Improving the locations further with hypoDD (Waldhauser and Ellsworth, 2000; Waldhauser, 2001) is the next stage that will hopefully shed more light on this.

To investigate the possiblitiy of missing foreshocks it would be best to carry out a similar study for future work using a different set of master events. The master events that we chose for this current project were to maximize the number of detected events. We found the ten best master events that could detect the largest percentage of the seismicity which happens to be aftershocks and background seismicity. Both the aftershocks and background seismicity occurred predominantly to the northwest of the 2004 Parkfield eartquake hypocenter. Therefore our newly detected events also occur close to the masters and far away from the M 6.0 hypocenter. There are a handful of events that are close to the mainshock hypocenter in the base catalog. These can be used as master events to see if they are able to detect any new foreshocks in the vicinity of the mainshock.

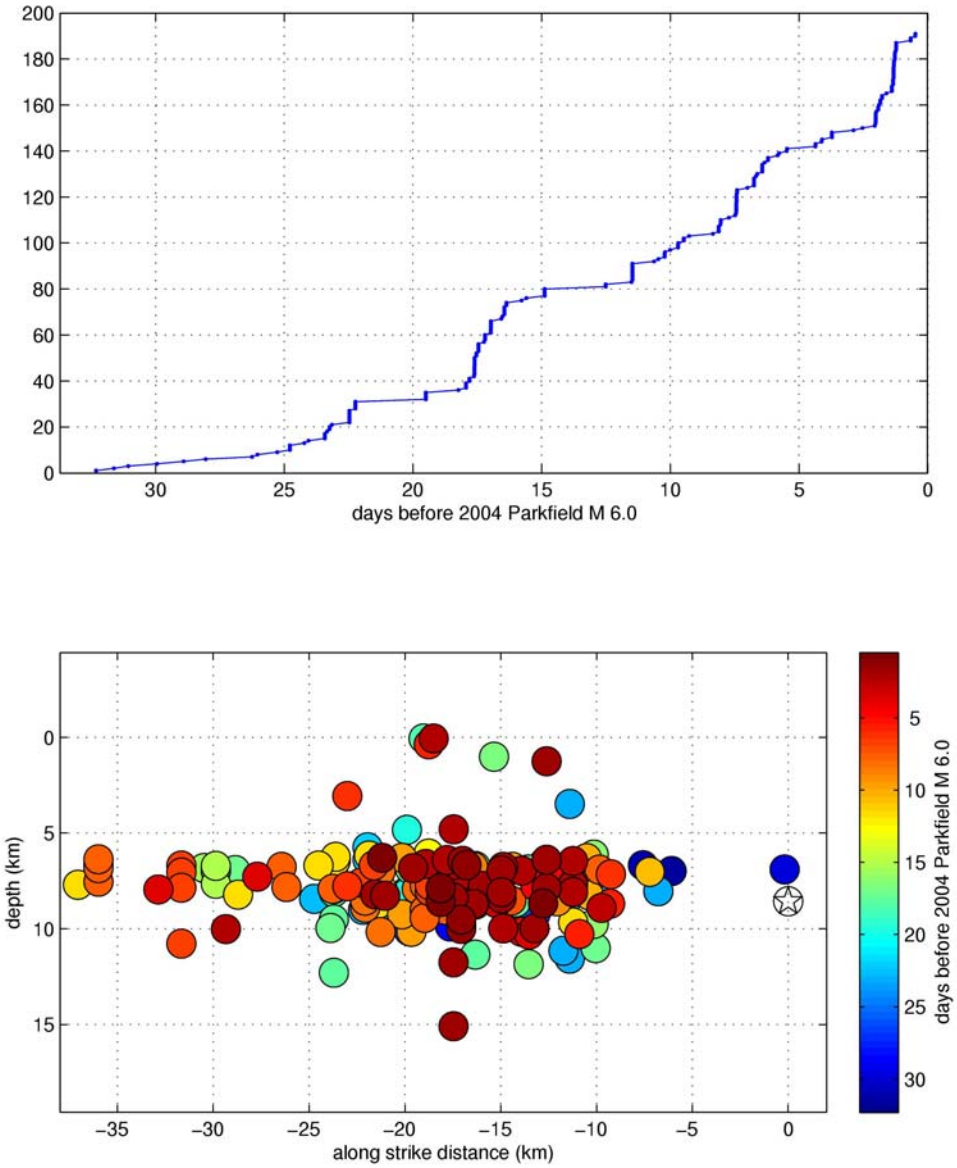

Figure 9. (top) Cumulative number of events in the new catalog of 6827 events in the days leading up to the 2004 Parkfield earthquake. (bottom) On-fault view of locations of events before Parkfield earthquake. Black star and circle indicates location of M 6.0 hypocenter of the mainshock.

3. References

Lengliné, O. and D.Marsan (2009). Inferring the coseismic and postseismic stress changes caused by the 2004 Mw = 6 Parkfield earthquake from variations of recurrence times of microearthquakes, *J. Geophys. Res.* 114, B10303, doi:10.1029/2008JB006118.

Peng, Z., J.E. Vidale, and H. Houston (2006). Anomalous early aftershock decay rate of the 2004 Mw 6.0 Parkfield, California, earthquake, *Geophys. Res. Lett.* 33, L17307, doi:10.10292006GL206744.

Peng, Z. and P. Zhao (2009). Migration of early aftershocks following the 2004 Parkfield earthquake, *Nature Geosc.* 2, 877-881, doi: 10.10138/NGEO697.

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, 5719-5713, 1984.

Saikia, C., 1994. Modified frequency-wave algorithm for regional seismograms using Filon's quadrature: modelling of Lg waves in eastern North America, *Geophys. J. Int.,***118**, 142-158.

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

Schaff, D. P., G.H.R. 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.

Schaff, D.P., Semi-empirical statistics of correlation detector performance, *Bull. Seismol. Soc. Am.,* 98, 1495-1507, doi: 10.1785/0120060263, 2008.

Schaff, D. P., Broad-scale applicability of correlation detectors to China seismicity, *Geophys. Res. Lett*., 36, L11301, doi:10.1029/2009GL038179, 2009.

Schaff, D., (2010) Improvements to detection capability by cross-correlating for similar events: a case study of the 1999 Xiuyan, China, sequence and synthetic sensitivity tests, *Geophys. J. Int., 180*, 829-846, doi: 10.111/j.1365-246X.2009.04446.x.

Schaff, D.P., and F. Waldhauser (2010), One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity, *Bull. Seism. Soc. Am., 100,* 3224-3238, doi: 10.1785/0120100042.

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. Seism. Soc. Am.* 96, S38-S49.

Van Trees, H.L., 1968, *Detection, Estimation and Modulation Theory,* John Wiley and Sons, New York.

Waldhauser, F., and W.L. Ellsworth, A double-difference earthquake location algorithm: method and application to the northern Hayward Fault, California, *Bull. Seism. Soc. Am., 90,* 1,353-1,368, 2000.

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

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

4. Papers/abstracts published related to this project

Schaff, D.P., and F. Waldhauser (2010), One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity, *Bull. Seism. Soc. Am., 100,* 3224-3238, doi: 10.1785/0120100042.

Schaff, D., (2010) Improvements to detection capability by cross-correlating for similar events: a case study of the 1999 Xiuyan, China, sequence and synthetic sensitivity tests, *Geophys. J. Int., 180*, 829-846, doi: 10.111/j.1365-246X.2009.04446.x.

Schaff, D.P., and F. Waldhauser, F. Broad-Scale Applicability of Correlation Detectors to China and Parkfield, California, Seismicity, Annual SSA meeting, Eastern section, Palisades, New York, October 2009.