IMPROVING MAGNITUDE DETECTION THRESHOLDS USING MULTI-STATION, MULTI-EVENT, AND MULTI-PHASE METHODS

David Schaff and Felix Waldhauser

Lamont-Doherty Earth Observatory, Columbia University

Sponsored by Air Force Research Laboratory

Contract No. FA8718-05-C-0022

ABSTRACT

Last year we reported on the results of both a semi-empirical analysis and a case study in Xiuyan, China, that demonstrated that a correlation detector can lower magnitude detection thresholds by over one full unit for similar events as compared to a standard STA/LTA detector. Ninety out of 90 events (100%) were detected by cross correlation for the case study whereas a procedure like the pIDC employs detected only 11%. We found that less than perfect matches still provided useful detections due to location, focal mechanism, and magnitude differences. Events with magnitude differences as large as 2.3 and 3.3 were shown to produce detection spikes on the correlation traces. Work is continuing applying the correlation techniques on a larger scale to 1,000 events at Parkfield, California, and 19,000 events in and near China. We are attempting to see how broadly applicable correlation methods can be applied to different tectonic settings and for what percentage of the seismicity. One hundred eleven million correlations were performed on Lg-waves for the events in China at 363 stations. Preliminary results suggest two thirds of the 19,000 events can be detected by cross correlation using this relatively sparse regional network. For Parkfield 82% of the events studied can be detected by cross correlation. Correlation detection is able to find additional events beyond what standard processing detects for China (70% increase) and for Parkfield (225% increase). Most event separation distances for events that correlate at Parkfield are less than 1 km. The distribution of magnitude differences for events that correlate at Parkfield is not distinguishable from the input magnitude distribution.

OBJECTIVES

Research will be conducted on techniques for generating a multiphase detection bulletin derived by two means: station array processing (multi-station) and source array processing (multi-event). Comparison and quantification of improvement over standard P-wave, single-event, single-station procession will be evaluated.

RESEARCH ACCOMPLISHED

Introduction

The second year of this project has focused on a correlation detector or multi-event technique applied to larger regions of seismicity. Waveform cross correlation has a long history of improving locations and identifying events (e.g., Poupinet et al. (1984); Harris (1991)). Recently it appears that large percentages of events may be similar enough to enable correlation detectors to be applied on a broad scale (Schaff and Richards (2004); Schaff and Waldhauser (2005)). The first year of the project demonstrated a full magnitude unit improvement in detection thresholds compared to a short-term average/long-term average (STA/LTA) type detector for similar events for both semi-empirical synthetic runs and a case study of 90 events in the 1999 Xiuyan earthquake sequence in China. The synthetic runs showed that this improvement is obtained with acceptable false alarm rates of about one per day. Gibbons and Ringdal (2006) also have shown order of magnitude improvement using a correlation detector for the Norwegian Seismic Array (NORSAR) array. The second year has applied these techniques to 18,886 events in and near China and to 539 events in Parkfield California.

Application to China

The left plot of Figure 1 shows the 18,886 events in and near China and 363 stations used for this study. The events come from the Annual Bulletin of Chinese Earthquakes (ABCE) from 1985 to 2005. The stations are those for which waveforms are available at the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC). Several of the stations are only temporary deployments. There are only a few long-running stations in China, which correlation techniques work best for, so the actual network of stations for most of the events is quite sparse with large inter-station distances. A total of 111 million cross correlations were performed taking about 2 weeks of continuous processing time on a four-CPU computer. All events with separation distances of less than 150 km were correlated. In other words, every event is a master event to see the maximum possible number of correlations possible. So far only Lg-phases have been processed. 50-s windows were used searching forward and 30-s windows searching backwards using time-domain cross correlation. The seismograms

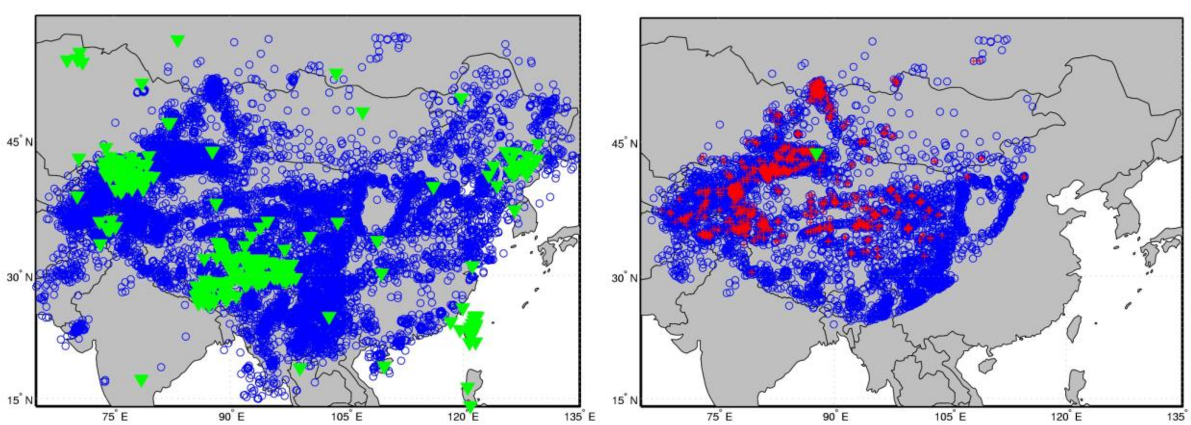

Figure 1. (left) 18,886 events (blue circles) recorded at 363 stations (green triangles) in and near China. (right) events in blue recorded at station WMQ (green triangle). 17% of the events (red) have CC > 0.5 with at least one other event at this station.

were filtered from 0.5 to 5 Hz. The cross correlation traces for the three components are averaged together to constructively enhance the detection spikes when present. A "scaled cross correlation coefficient" (SCC) was used to initially sift the data. All correlations with SCC > 4.5 were saved.

A study was made to estimate false alarms using real data for selected stations. It is assumed that randomly selected time windows should not contain events that correlate. From this it is possible to determine curves of the false alarms per day as a function of cross correlation coefficient (CC) and SCC. For traces that are already sifted with an SCC 6, a CC of 0.24 corresponds to approximately one false alarm per day. A CC of 0.5 corresponds to 0.0022 false alarms per day. An SCC of 6.65 corresponds to a little over one false alarm per day. Based on the pair-wise distance matrix for the correlations and the number of samples in the lags searched over, we can estimate the percent of the 18,886 events for which detections are processed that would be expected to be false alarms. Because a CC of 0.5 appears to be so robust for these long 50 s windows we assume that a trigger at a single station provides a reliable detection. The right plot of Figure 1 shows the events within 20 degrees recorded at station WMQ. The events in red are 17% of the total that have CC > 0.5. Notice how the Tibetan Plateau blocks the Lg-wave propagation for this station. For the other thresholds of one false alarm per day, it is necessary to require at least two stations observing that pair to count as a detection. Using a combination of these criteria (CC > 0.5 at 1+ stations, CC > 0.24 at 2+ stations, or SCC > 6.65 at 2+ stations) we estimate the percent of events that would be false alarms to be approximately 3%. Applying these criteria to the time windows corresponding to where the theoretical travel times would arrive for the 18,886 events results in 12,902 events that are detected or 68%. Therefore it is estimated that 65% represent true detections.

Figure 2 shows the magnitude distribution for the 18,886 events and the 12,902 events detected by cross correlation. Eight thousand, three hundred fifty-eight events found by a "pIDC" type detector are also shown (44%). The pIDC employed an STA/LTA detector on the vertical component for P-waves in overlapping narrow pass bands. Three stations triggered are necessary to be counted as a detection. The number in parentheses in the legend is the 95% confidence lower limit of the magnitudes detected. Overall it can be seen there is a 0.2 unit reduction in magnitude detection threshold for correlation compared to the "pIDC". However 2.8 is also the 95% lower limit of all the magnitudes in the catalog. Therefore the range of magnitudes is not low enough to test if correlation improves detection thresholds more than 0.2 units overall.

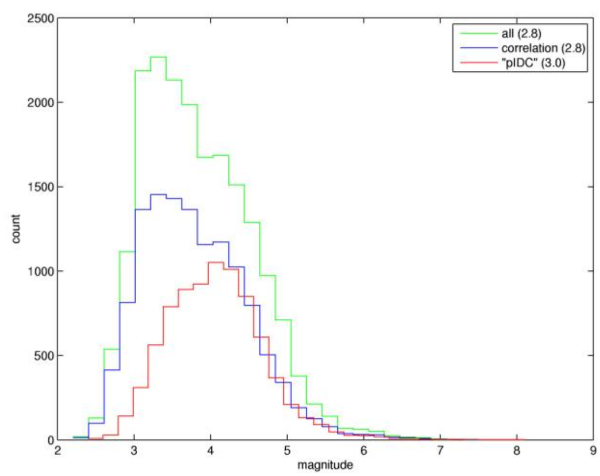

Figure 2. Histograms of the magnitude distribution for all 18,886 events and the 12,902 events detected by correlation and 8,358 found with pIDC type procedures. Correlation finds more events and lower magnitude thresholds restricted by the lower limit of the catalog magnitude of completeness. The number in parenthesis in the legend gives the 95% confidence lower limit for the magnitude distribution.

More insight can be gained by plotting magnitude as a function of station distance. The left plot of Figure 3 shows the detections for the “pIDC”. The red line is the 95% lower limit of the magnitudes in 50 km bins of station distance. The trend is seen to increase from about 2.9 at zero to 3.5 for 2,200 km station distances. The right plot of Figure 3 shows that the number of observations that are detected steadily decreases after about 500 km. Figure 4 shows similar plots for the correlation detector. The red line is the 95% lower limit of the magnitudes in 50 km bins. The green line is the 95% trend from Figure 3. It is seen that the two lines diverge for greater station distances. The right plot of Figure 4 shows that correlation detection observations drop off faster with station distance. The difference between the two 95% trends is plotted in Figure 5. At zero the difference is 0.2 units and then increases to a maximum of 0.9 units at greater station distances. The interpretation of these results is that for longer station distances the magnitudes in the catalog are sufficiently complete to observe nearly an order of magnitude improvement in threshold reduction between the two techniques. For closer station distances lower magnitude events are not available in this catalog to test if the full unit reduction still holds.

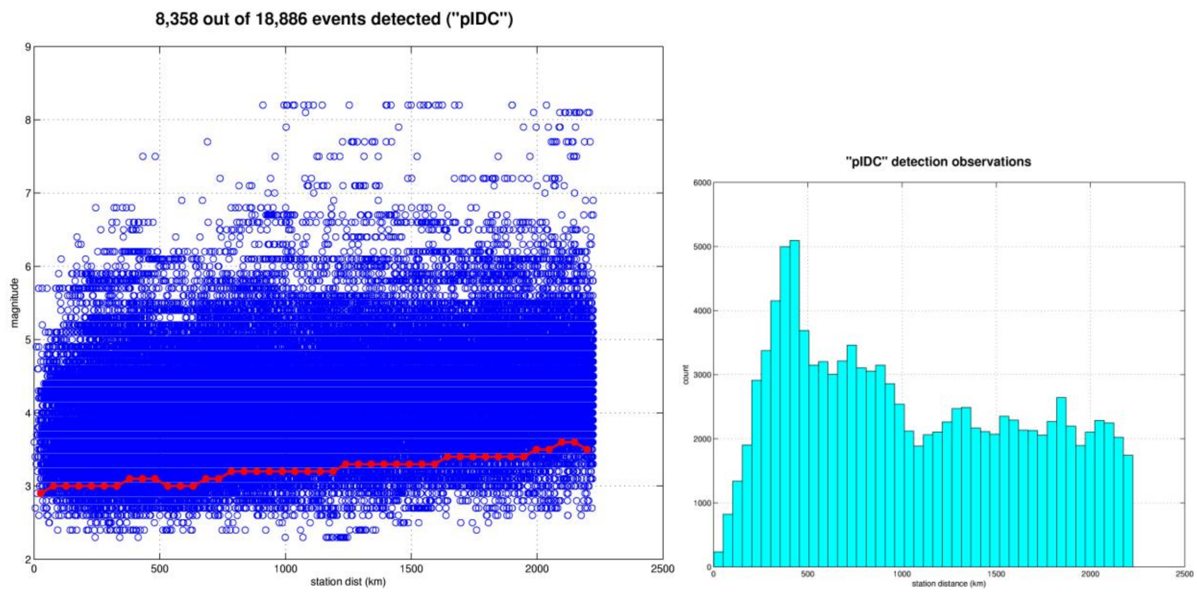

Figure 3. (left) Plot of magnitude vs station distance for all triggers for “pIDC”. Red line shows 95% confidence lower limit of magnitude in 50 km bins. (right) Histogram of number of observations as a function of station distance.

We see that approximately two thirds of the events in the catalog are detected by cross correlation (12,902), a sizeable fraction indicating that these methods can be applied on a broad scale across diverse tectonic regions. However, this catalog was based on a much denser network of stations than the one that we have waveforms available for. To get an idea of how applying correlation methods to an existing network would improve things it is instructive to see how many events are detected by correlation and an STA/LTA detector on the same network. Seven thousand, sicty-three events were detected by correlation out of the 8,358 found by a “pIDC” procedure or 85%. For comparison Schaff and Waldhauser (2005) determined that 95% of the 225,000 events in northern California correlated at four or more stations with CC > 0.7 with at least one other event. Therefore correlation is able to detect the great majority of the seismicity for these large regions of seismicity. This can be an important independent confirmation of the existence of new events and help to weed out false alarms. Besides lowering magnitude detection thresholds correlation also detects more events that the “pIDC” procedure missed due to a variety of reasons (Figure 1, right plot). The correlation detector finds 5,839 additional events over the 8,358 events from the “pIDC” detector or a 70% increase. Therefore we might expect catalogs for existing networks to also increase with the complementary benefits of correlation detector techniques.

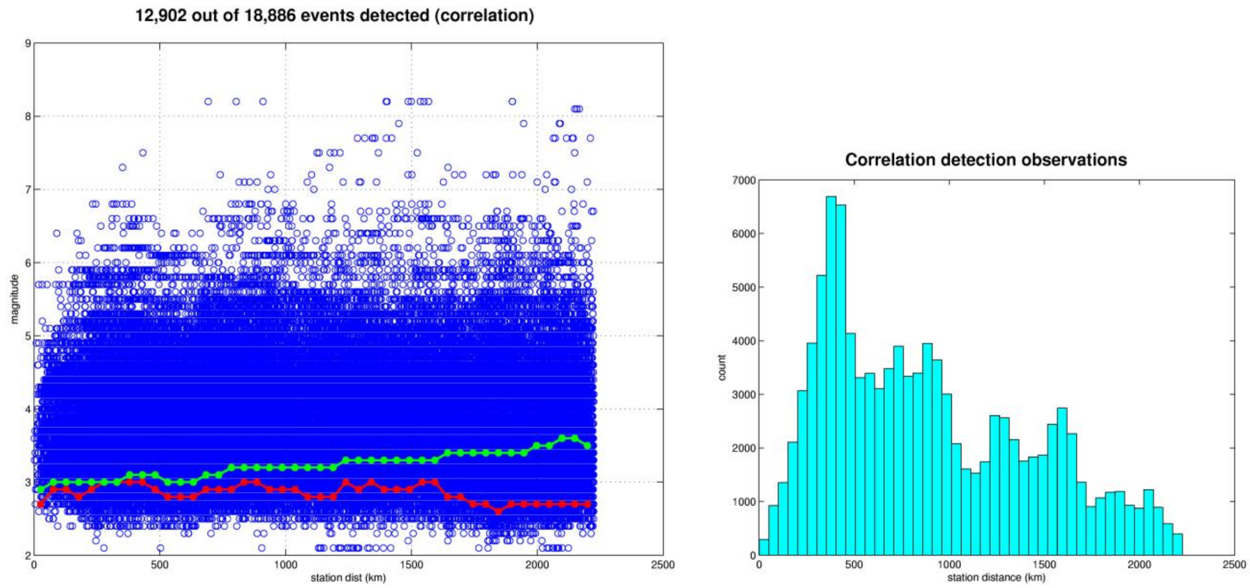

Figure 4. (left) Plot of magnitude vs. station distance for all triggers for correlation detector. Red line shows a95% confidence lower limit of magnitude in 50 km bins. Green line is the curve for the "pIDC" from Figure 3. (right) Histogram of number of observations as a function of station distance.

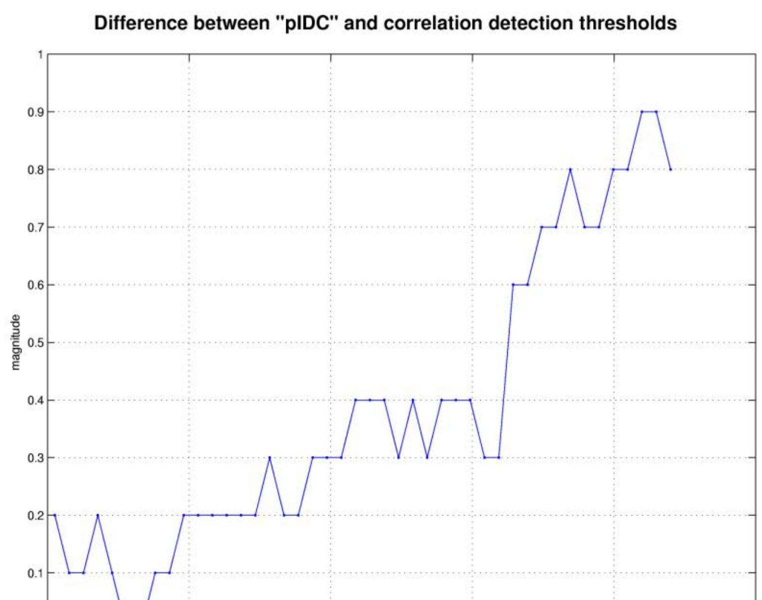

Figure 5. Difference between "pIDC" (green) and correlation (red) lines on Figure 4 as a function of station distance.

Application to Parkfield, California

Results are preliminary for Parkfield. Five hundred thirty-nine events with waveforms available and magnitudes of 2 or greater are analyzed at five continuous stations archived at IRIS (Figure 6). The first arriving P-wave, S-wave, and Lg-wave are processed. The window lengths are 15 s, 20 s, and 50 s respectively. The lags searched over are 10 s, 15 s, and 30 s respectively. The filter bands are 0.75 to 2 Hz, 0.5 to 3 Hz, and 0.5 to 5 Hz respectively.

In this case because of the limited number of stations we are able to estimate false alarms directly for each station. The codes are run as normal except the windows selected are 120 s before the first arriving P-wave which is assumed to contain only noise. The window lengths, lags, and filter bands are kept the same for each "phase". Based on these results we are able to determine that only 8 out of the 539 events or 1.5% would be expected to trigger as false alarms using a criteria of SCC > 6.3 at two or more stations or phases. Applying this same criteria to the normal processing for windows centered on the theoretical arrival times of the phases for the 539 events yields 444 events that are detected by cross correlation or 82% of the M 2 or larger events in the study area. A cross section of these 444 events is shown in Figure 7.

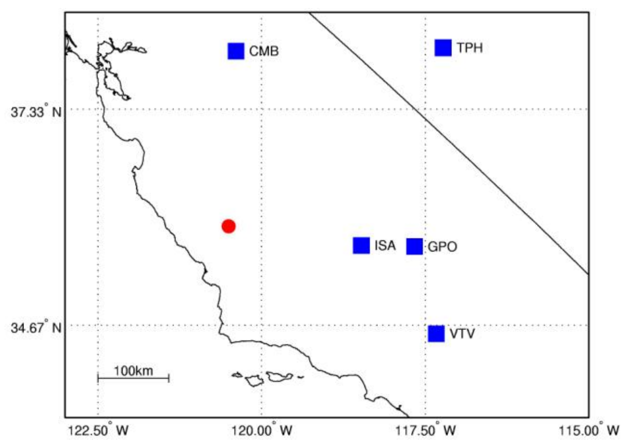

Figure 6. Map of California showing location of Parkfield study area (red dot) and five continuous stations used (blue squares).

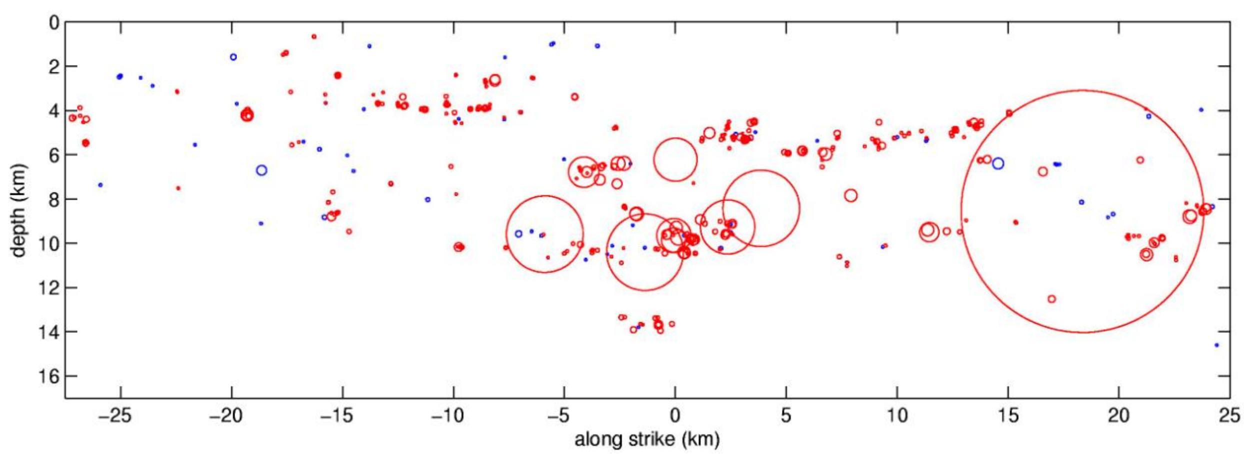

Figure 7. Along fault cross section from NW to SE for 539 events studied that have M greater than or equal to 2. Four hundred forty-four events (82%) in red are detected by cross correlation. Blue events are missed detections. Source areas for events are estimated based on a circular crack model and a 3 MPa stress drop.

The distribution of magnitudes for the correlation detections and “pIDC” is plotted in Figure 8. Because of the limited number of stations considered for the “pIDC” detector we require only 2 stations to trigger to count as a detection instead of 3. It is seen that the correlation detector finds 311 additional events over the 138 found from the “pIDC” or a 225% increase. In this case for these station distances both detectors are able to find M 2 events at the 95% confidence limit so no reduction in magnitude threshold is observed. Fortunately the catalog in this region extends to magnitudes of 1 and 0 so we can most likely test if similar order of magnitude improvements are observed as seen for China and at NORSAR. If the increase in number of events detected is an indication we might expect this general rule to also hold. Out of the 138 events that are found by the “pIDC” detector 133 events are detected by cross correlation or 96%.

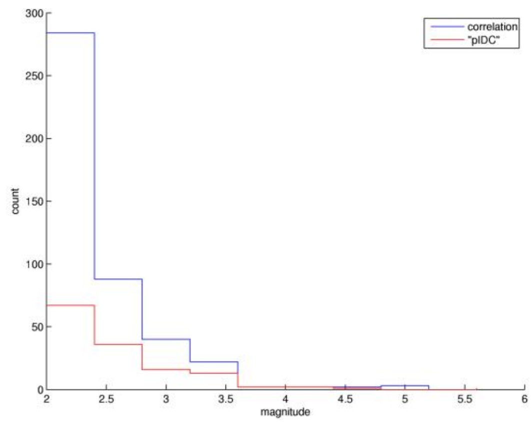

Figure 8. Histogram of the distribution of magnitudes detected by correlation and “pIDC” for Parkfield.

The left plot of Figure 9 shows the distribution of event separation distances for detections that met the criteria of SCC > 6.3 at two or more stations or phases. It is observed that they strongly cluster less than one kilometer with a median of 0.8 km. The pair-wise input correlations considered were all events separated by 20 km or less. A distribution of that observation matrix is shown in the right panel of Figure 9 with a median of 9 km. The difference between the two distributions is striking confirming the expectation that closely separated events correlate the best and that there are few false alarms in the results as estimated (1.5%).

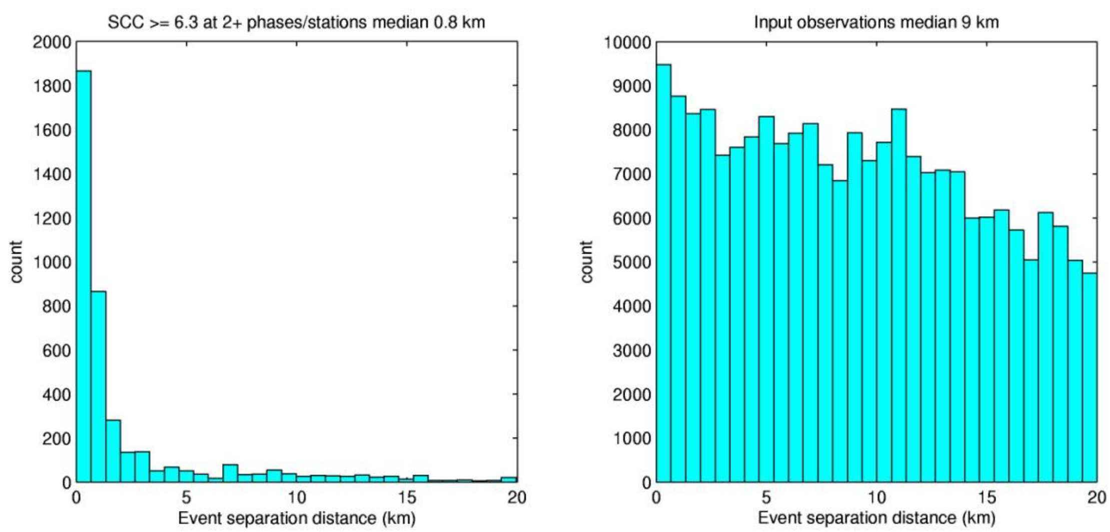

Figure 9. Histograms of event separation distance of correlation detections (left) and input matrix (right).

Similar plots are shown in Figure 10 except for the magnitude difference between the pairs. It is interesting to note that this time the median magnitude difference and distribution is similar for the two comparisons suggesting that events with greater difference in magnitude can correlate well enough to provide useful detections. This is helpful to know because one of the applications we would like a correlation detector to be able to do is to use a larger magnitude event to hopefully detect a smaller magnitude event that is buried in the noise. These results are consistent for a large set of events with examples shown for the Xiuyan case study of event pairs with 2.3 and 3.3 magnitude unit differences that correlated well enough to be detected.

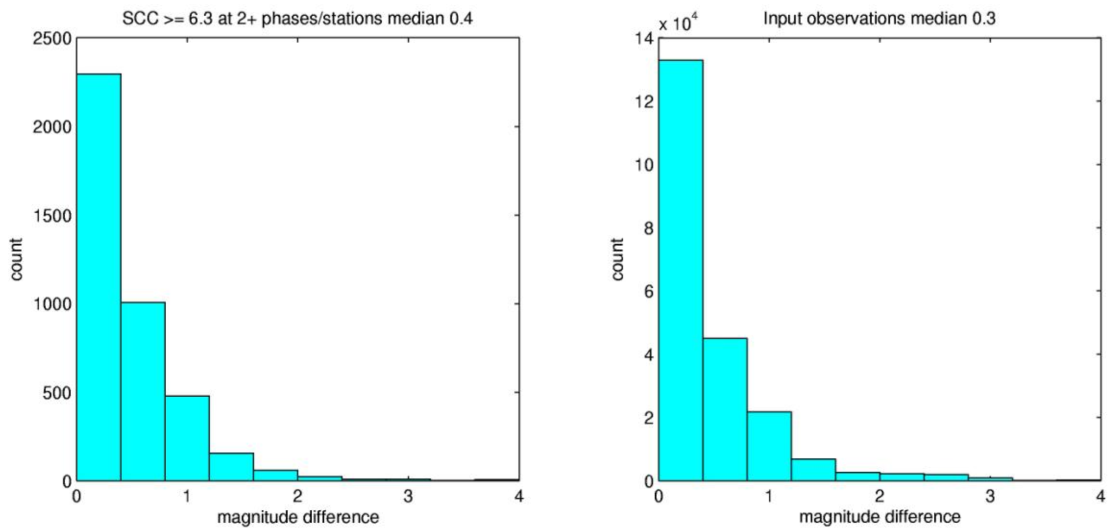

Figure 10. Histograms of magnitude difference of correlation detections (left) and input matrix (right).

CONCLUSIONS AND RECOMMENDATIONS

- Correlation detects a sizeable fraction of the seismicity for China (about two thirds) and for Parkfield (82%) with false alarm rates of a few percent.
- When correlation detectors and STA/LTA filters are applied to the same network, correlation finds the significant majority of detections that standard processing detects for China (85%), for Parkfield (96%), and for northern California (95%).
- Correlation provides an independent and complementary means to confirm detections to STA/LTA and can help to weed out false alarms.
- Correlation detects additional events beyond standard processing for China (70% increase) and for Parkfield (225% increase).
- Overall for China a reduction in magnitude detection threshold of 0.2 units is observed for correlation compared to “pIDC” procedures. Closer examination reveals that this difference increases to 0.9 units with increasing station distance. It is assumed that magnitude completeness of the catalog is the reason a larger improvement is not seen for closer station distances. Lower magnitudes and further station distances need to be examined for Parkfield to see if there is any reduction in magnitude detection threshold.
- Most event separations that correlate well enough to trigger a detection at Parkfield are less than 1 km even though all event separations out to 20 km were searched.

- Magnitude differences between the pairs of events that correlate have a distribution that is virtually indistinguishable from the input distribution indicating that less than perfect waveform matches still provide useful detections.

ACKNOWLEDGEMENTS

We are grateful to the operators of the Chinese Digital Seismic Network and to the IRIS data management center for archiving the waveforms used in this project. We thank Paul Richards, Won-Young Kim, and Bill Menke for helpful discussions. We thank Bob Bookbinder and Dave Lentrichia for computer support.

REFERENCES

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* 165: 149–166.

Harris, D. B. (1991). A waveform correlation method for identifying quarry explosions, *Bull. Seism. Soc. Am.* 81: 2395–2418.

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

Schaff, D. P., and P. G. Richards (2004). Repeating seismic events in China, *Science* 303: 1176–1178.

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