Automated Techniques for Waveform Correlation Applied to Regional Monitoring of Eastern Asia

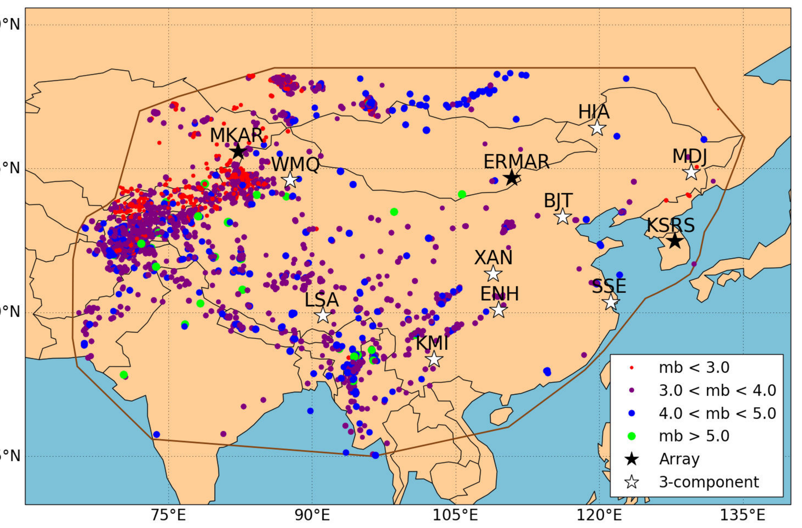
Amy Sundermier[1*], Megan Slinkard[1], James Perry[1], David Schaff[2], Christopher Young[1], Paul Richards[2] | *Sandia National Laboratories*[1], *Lamont-Doherty Earth Observatory*[2]

T3.7-P4

INTRODUCTION

Waveform correlation techniques have proven effective detecting repeated events from large aftershock sequences; however, application for monitoring a large region over a long time period has yet to be adequately explored. We applied waveform correlation to continuous waveform data for the year of 2012 at twelve stations spread through Eastern Asia, using automatically generated templates from historical archives going back to the time of station installation. Our study region includes the countries of China, North Korea, South Korea, Mongolia, Nepal, Bhutan, Bangladesh, and parts of neighboring countries. We used nine China Digital Network (CD/IC) three-component stations and three arrays that had continuous coverage from 2006-2012; this yielded 12 stations which spanned 40 degrees in latitude and 70 degrees in longitude with a mean nearest-neighbor distance between stations of 728 km. To declare a detected event, we required coincident correlations from two or more arrivals. Detection results are compared to the International Seismological Centre (ISC) catalog to analyze the effectiveness and challenges associated with applying waveform correlation on a broad regional and multi-year scale. We analyze the impact of network geometry, historical template library span and size, and template phase to provide direction for future regional studies using waveform correlation.

STATION SELECTION AND EXPERIMENT SETUP



Study period	1/1/2012 – 12/31/2012 (continuous data processed)
Study area	12 stations in Eastern Asia
Template events	ISC events from 1986 - 2011

The map shows the ISC Events in 2012 by body wave magnitude that we want to detect with template waveforms from events prior to 2012. Stations are indicated by stars. The polygon delimits the study area.

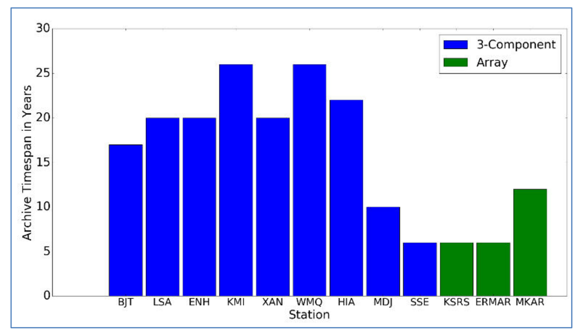
TEMPLATE PREPARATION

A template library is a collection of waveforms from known events used to search for similar events seen at a station. Template waveforms should be clear signals with a high signal-to-noise ratio. We explored approaches to automating the template building process.

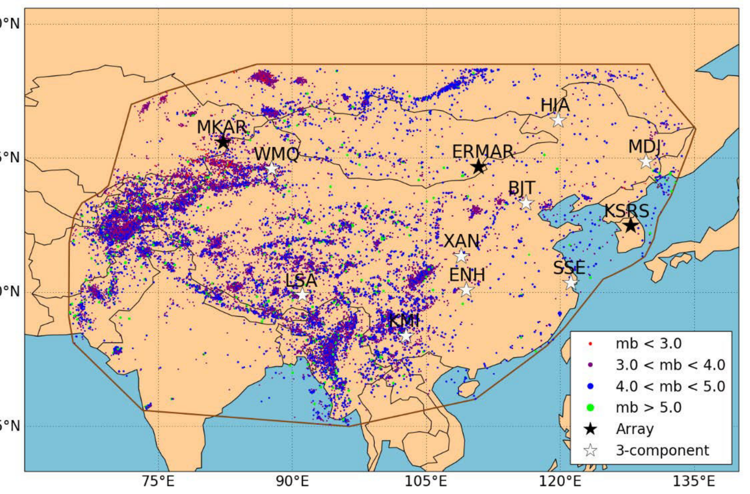
Template Preparation Process

Query event origins in ISC bulletin, and calculate arrival times for events. → Filter waveforms (parameter table) → Use STA/LTA to verify observable arrival at station → Select correlation threshold (template-specific) → Template library

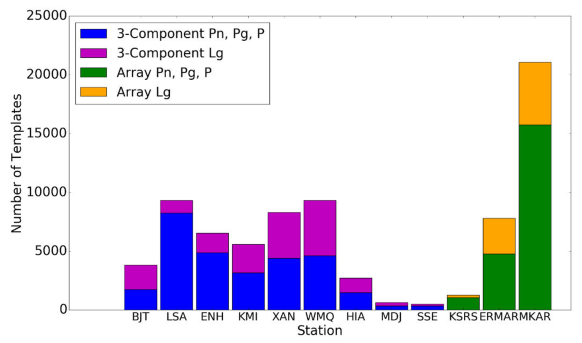
Parameters for Template Selection and Screening		
Phase	First P	Lg
Arrival Prediction Method	AK135 model / Tau P	3.5 km/s group velocity
Template Window (sec)	25 (-5/20)	30 (-5/25)
Filter (Butterworth 3-pole) (Hz)	0.8-3.0	0.5-5.0
Propagation radius	Not applicable	20 degrees
STA/LTA window and gap (sec)	1/10/0	3/30/100
STA/LTA Threshold	3.0	3.0



The temporal extent of recorded data used for template generation and waveform correlation. We used up to 25 years of historical archives.



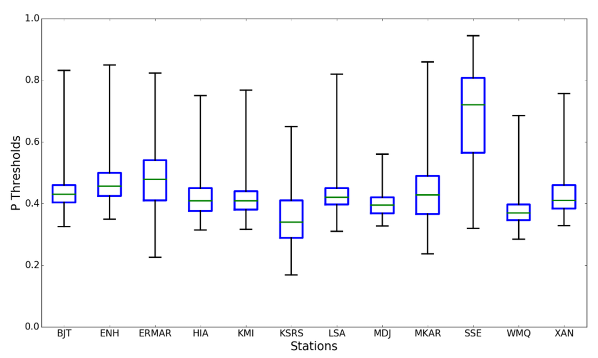
The map (left) shows the template waveforms by body wave magnitude, created from station historical archive waveforms through 2011. The stations are indicated by stars.

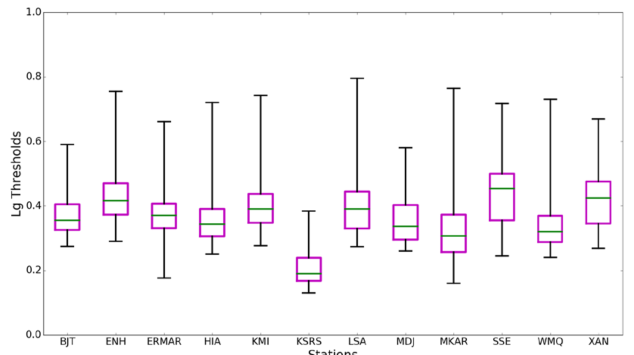


The bar graph (right) shows the number of templates by station and phase.

Template Threshold Selection

Using the reverse template method (Slinkard et al, 2014), we automated the process for identifying a correlation threshold for each template. Thresholds were set for a consistent FAR or 3/yr across the template library.



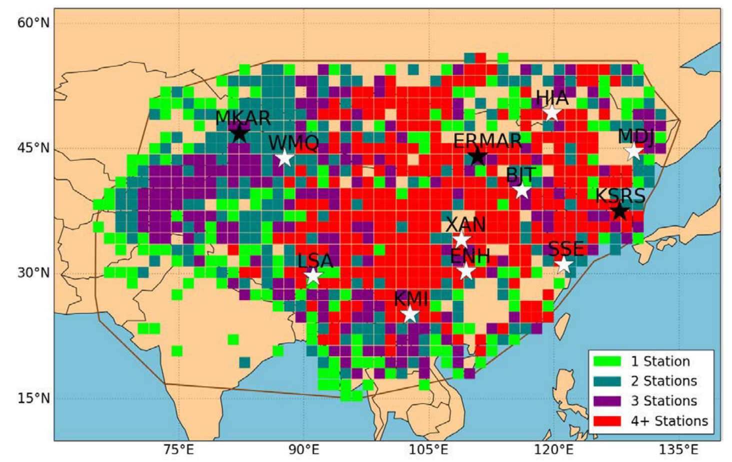


Statistics for the P template thresholds by station (left). Statistics for the Lg template thresholds by station (middle). In both diagrams the box extends from the lower to upper quartile of the thresholds with a line representing the median threshold. The minimum (maximum) threshold is the bottom (top) line of the distribution.

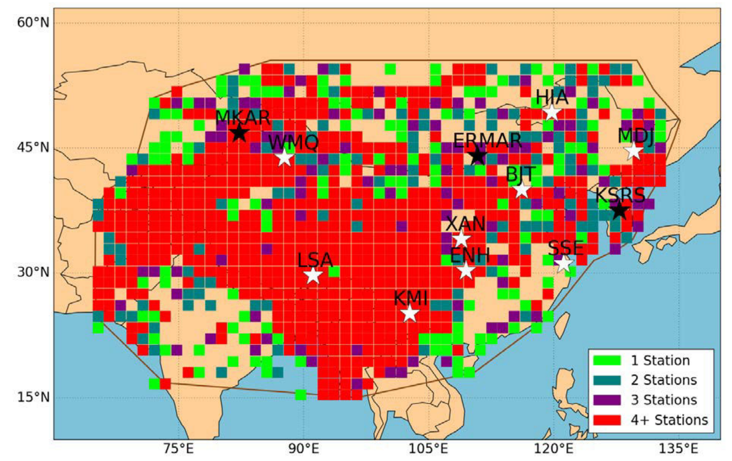
Note the wide range of thresholds selected within each station's templates; this demonstrates the importance of using individualized thresholds. Templates prone to false detections are assigned high thresholds, templates not prone to false alarms have lower thresholds. Across the collection of stations, the median Lg template threshold is lower than the median P template threshold. This is not surprising given the Lg templates generally have a higher time-bandwidth product than templates from Pn, Pg, or teleseismic P phases.

SPATIAL GRID OF THE TEMPLATES

The maps plot a spatial grid of the number of templates per 150 km grid square. Having multiple stations with templates in the same geographical area increases our ability to validate detections.



Lg The coverage of Lg templates is highest in the areas of densest station locations. We restricted template formation for Lg phases to events within 20-degree radius of a station.



First P The coverage of P templates is more complete for the highly active seismic areas in the western region of the study area.

The spatial distribution of the templates suggests that we should detect much of the seismicity for 2012 from the region, but if the source for the template is not similar enough to the source for an event within the continuous data stream then the waveforms will not correlate above the template detection threshold; thus the grid coverage of 4+ templates does not guarantee detection.

WAVEFORM CORRELATION PROCESSING

Waveform correlation processing consists of correlating each template with the raw data stream at each station, and declaring a single-station detection whenever the correlation value exceeds the correlation threshold. This is computationally expensive, so our SeisCorr software distributes the processing on a cluster.

Correlate historical templates with continuous waveform data for each station. → Identify detections (correlation score > template threshold). → Estimate location and magnitude (based on template event data)

Correlation Statistics				
Station Type	Number of Stations	Channels	Templates	Processing (Hours) for 1 year
Three-Component	9	27	47,840	161
Array	3	37	30,204	392

MULTISTATION and MULTIPHASE VALIDATION

Multistation validation is a post-processing step where detections from different stations are compared, and detections are validated. Recall we anticipate 3FA/yr per template at each station, so, we anticipate many false detections during a year. However the probability of more than one station detecting a false detection at the same time is small, so we require more than one template detect each event. This gives us high confidence in our validated detections, however, multistation validation limits our ability to search for small events with a sparse regional network, as many events are simply not seen at more than one station.

Multiphase validation occurs when templates from multiple phases (e.g. the P and Lg templates) at the same station detect within the selected tolerance for an event. This allows us to use the same principal of reducing the rate of false alarm, but allows events seen at only 1 station to be validated.

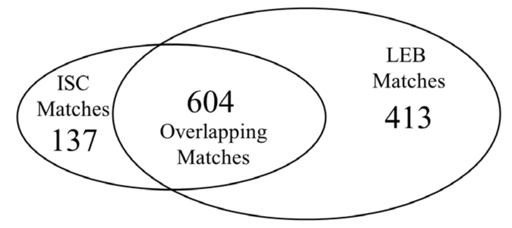
Compare single-station detections to find detections with calculated origin times within 50 sec. → Verify that finding-templates were geographically collocated within 150 km (e.g. ensure detections not from events that occurred at the same time on opposite sides of the world) → Validated Detections

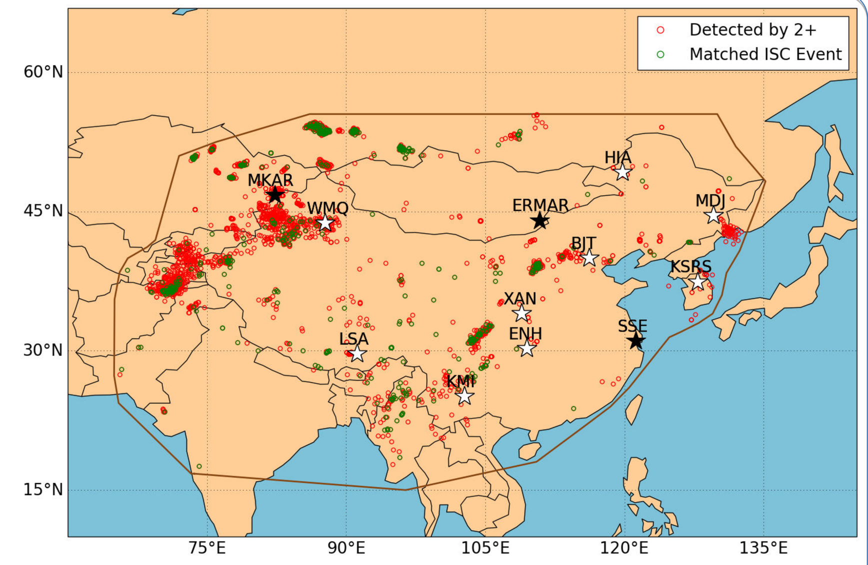
We had 6130 validated events during our study period of 2012. Just over half of these (51%) were from events seen at only one station, validated using multiphase validation.

RESULTS

Summary of Detected Events and Comparison with Bulletins			
	2012 Events in Catalog	6130 Events from 2+ Stations/Phases Detections	
		Found Catalog Event	Additional Events
ISC Bulletin	3597	741 (21%)	5389
IMS Late Event Bulletin	5138	1017 (20%)	5113

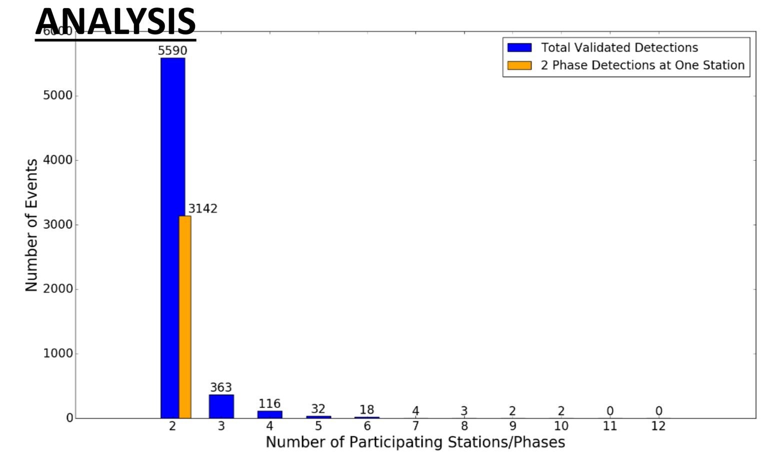
Venn diagram of the number of detected events that matched the ISC bulletin, the LEB, and both bulletins.



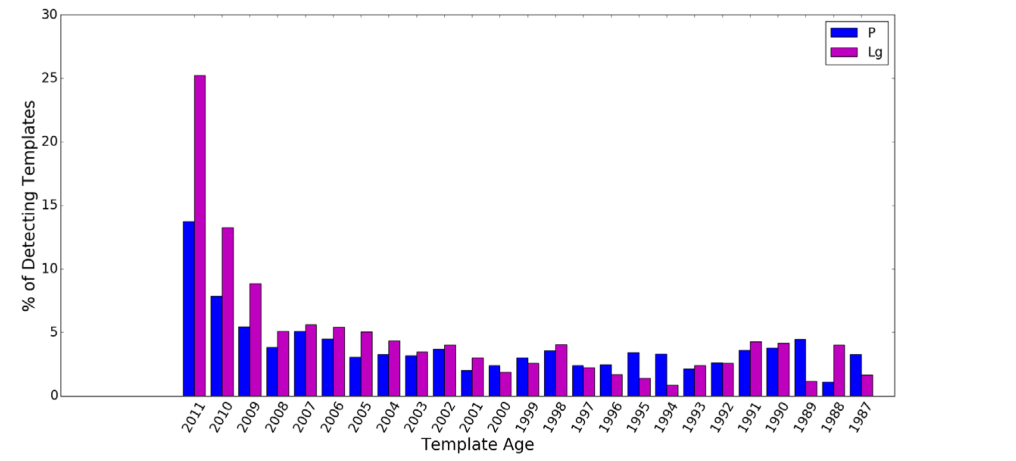


Map of the multistation validated events, with the 741 events that matched the ISC bulletin as green circles. The additional 5389 events that were detected by waveform correlation and validated may be legitimate seismic events but we have no independent way to verify them.

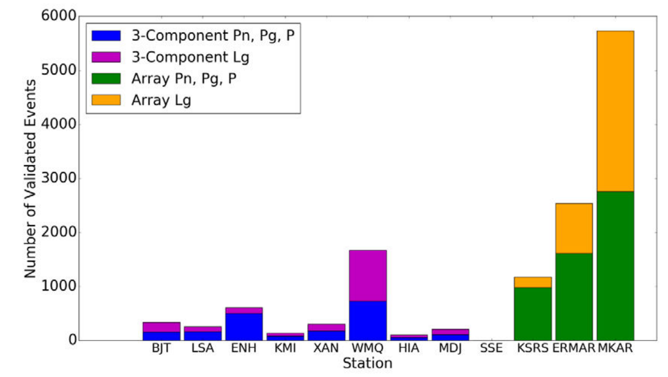
ANALYSIS



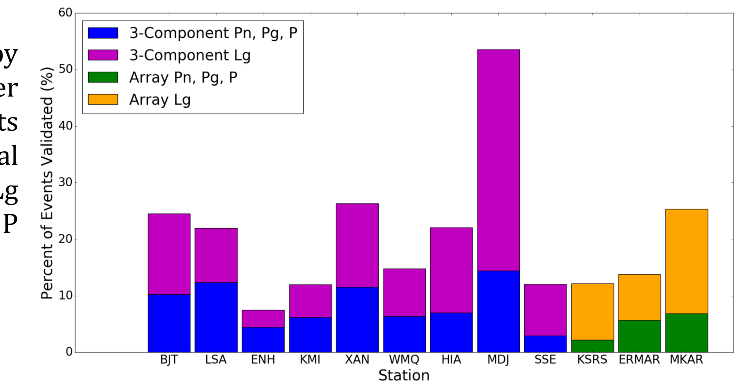
Above graph illustrates the number of events that were detected by the given number of stations/phases for 2012. Network geometry resulted in many events being too far from 3 or more stations to be validated using multistation validation.

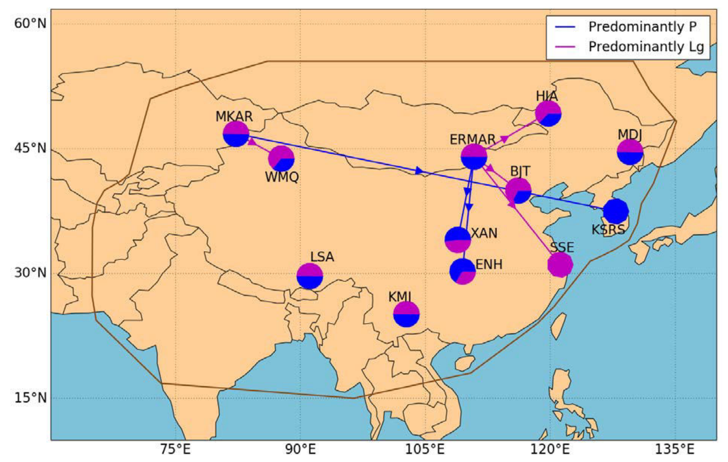


Above graph illustrates the ages of the templates that detected events in 2012, normalized by the number of templates from the origin year; few templates were available prior to 1995. The year that generated the most templates was 2008; likely due to the Wenchuan aftershocks. We speculate that recent templates from mine explosions account for the success of templates created in 2011-2009, while detections by tectonic templates is more even throughout the graph. This graph indicates the importance of using a long historical archive to detect tectonic activity by waveform correlation.



Number (left) of events that were detected by phase at each station and validated by another station/phase. The arrays detected more events than the 3C stations despite their shorter historical archives. The percent (right) illustrates that Lg detections are validated more often than P detections for both arrays and 3C stations.



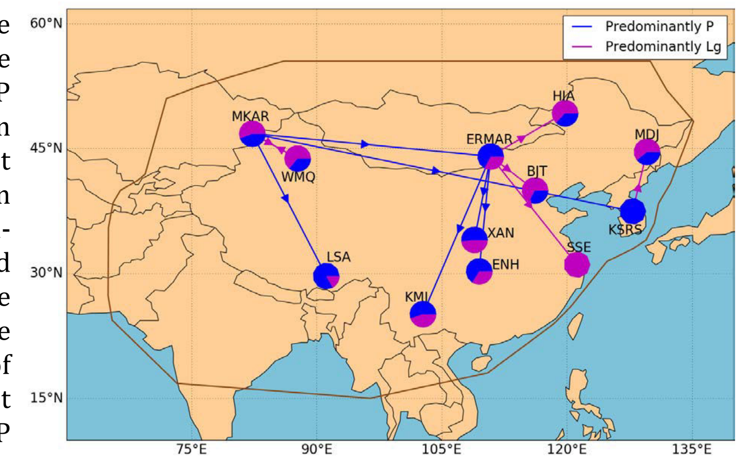


Map of the best validating station for every station, where a station detection may *self-validate* if the same event is detected with templates from two different phases. Isolated pie charts indicate stations that are self-validating; detecting two independent phase arrivals at the same station. ERMAR, KMI, LSA, MDJ and MKAR validated their own detections more often than other stations validated their detections.

For the network view of our results, the pie chart at the station location represents the ratio of detections from the Lg vs. first P templates that were validated at that station by the best validating station. The best validating station detects the most events in common with a specific station. For station-pairs, the best validating station is indicated by the wide end of the arrow, while the station being validated is indicated by the point of the arrow on the line. The color of the line and arrow show whether the best validating station confirmed more Lg or P template detections.

Analyzing the pairs of stations that detected the same events revealed that there are three major factors that predict a station's performance in our study:

1) station was an array,
2) station is close to the seismicity,
3) station is close to other stations that are within detection range of the seismicity.



Map of the best validating other station for every station. Array ERMAR was the most frequent best validating station for six of twelve stations in the network; its central location, proximity to three-component stations with Lg propagation range, and superior sensitivity resulted in coincident detections. The analysis also shows that the ratio of first P/Lg validations is strongly correlated to the interstation distance for the station-pairs.

SUMMARY

This research was a practical test of the effort needed to assemble waveform templates, run cross-correlation over continuous data with a predetermined false alarm rate, and build new events over a broad area. Our study reviewed the impact of network geometry, historical waveform template library span and size, and template phase to provide directions for incorporating automated waveform correlation techniques into modern monitoring processing schemes. Analysis of our results found that using templates from both P and Lg arrivals improved detection clustering and enabled validating of events seen at only one station. Templates from the Lg phase were most valuable for detections, however, the P templates are important for validation and overall network performance; in particular, teleseismic P is useful for validation from distant stations. The arrays outperformed the three component stations both in the number of events detected and pairing with other stations in the network to validate detections due to their superior sensitivity. We detected 20% of the ISC Bulletin events and over 5000 additional events. Future work may include improvements to the automated template generation process, analysis of multi-template detections, and research for reducing false alarms for single station events.

Slinkard, M., D. Schaff, N. Mikhailova, S. Heck, C. Young, and P. G. Richards (2014). Multistation validation of waveform correlation techniques as applied to broad regional monitoring, *BSSA*. **104**, no. 6, 2768-2781, doi: 10.1785/01020140140.
Slinkard, M., S. Heck, D. Schaff, N. Bonal, D. Daily, C. Young, and P. G. Richards (2016). Detection of the Wenchuan Aftershock Sequence Using Waveform Correlation with a Composite Regional Network, *BSSA*. **106**, no. 4, 1371-1379, doi: 10.1785/0120150333.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This work was also supported by the Defense Threat Reduction Agency under Award Number HDTRA-1-11-1-00027 to Columbia University. The views expressed here do not necessarily reflect the views of the United States Government, the United States Department of Energy, the National Nuclear Security Administration, the Defense Threat Reduction Agency, or Sandia National Laboratories.

asunder@sandia.gov // Ground-Based Nuclear Explosion Monitoring Research and Development

SAND2017-5302 A