Bulletin of the Seismological Society of America, Vol. 104, No. 6, pp. 2768–2781, December 2014, doi: 10.1785/0120140140

Multistation Validation of Waveform Correlation Techniques as Applied to Broad Regional Monitoring

by Megan Slinkard, David Schaff, Natalya Mikhailova, Stephen Heck, Christopher Young, and Paul G. Richards

Abstract Waveform correlation is garnering attention as a method for detecting, locating, and characterizing similar seismic events. To explore the opportunities for using waveform correlation in broad regional monitoring, we applied the technique to a large region of central Asia over a three-year period, monitoring for events at regional distances using three high-quality stations. We discuss methods for choosing quality templates and introduce a method for choosing correlation detection thresholds, tailored for each template, for a desired false alarm rate. Our SeisCorr software found more than 10,000 detections during the three-year period using almost 2000 templates. We discuss and evaluate three methods of confirming detections: bulletin confirmation, high correlation with a template, and multistation validation. At each station, 65%–75% of our detections could be confirmed, most by multistation validation. We confirmed over 6500 unique detections. For monitoring applications, it is of interest that a significant portion of the Comprehensive Nuclear-Test-Ban Treaty Organization's Late Event Bulletin (LEB) catalog events was detected and that adding our confirmed detections for the LEB catalog would more than double the catalog size. Waveform correlation also allows for relative magnitude calculation, and we explore the magnitudes of detected events. The results of our study suggest that doing broad regional monitoring using historical and real-time-generated templates is feasible and will increase detection capabilities.

Introduction

Waveform correlation has many strengths as a method for detecting, locating, and characterizing similar seismic events. It has been shown to find events up to one order of magnitude smaller than can be conventionally detected (Schaff, 2009, 2010; Schaff and Waldhauser, 2010), find events buried in noise (Schaff, 2008), offer easy extension to double difference and other relative relocation methods to improve relative location estimates (Hauksson and Shearer, 2005; Shearer *et al.*, 2005; Waldhauser and Schaff, 2008), and can allow characterization of an event as an earthquake, mining event, or nuclear explosion from a signal recorded at only one station (Gibbons and Ringdal, 2012; Schaff *et al.*, 2012). Previous work has suggested that a significant portion of worldwide seismicity is from similar events that waveform correlation might be well suited to detect (Schaff, 2009; Schaff and Richards, 2011; Slinkard *et al.*, 2013; D. Dodge, personal comm., 2013). These results suggest that waveform correlation might have broad application to monitoring of earthquakes and nuclear explosions, events for which detection and characterization of seismic events are desired to be performed ever more quickly, accurately, and at lower magnitudes. If this conclusion is true, there are many implications for network operations and bulletin preparation. For example, using waveform correlation, a sparse network of high-quality stations that has a long continuous waveform archive can potentially provide the data for a seismicity bulletin with precise locations and include events down to lower magnitudes than are usually documented. The obvious caveat is that cross-correlation methods will not help to detect or analyze events dissimilar to any previously recorded events, but over time such events may become increasingly rare.

To explore several issues related to routine application of waveform correlation for seismic monitoring, we applied the technique to a broad region of central Asia over a three-year period at three high-quality stations at regional distances. This work relied upon the development by Sandia National Laboratories of a high-performance distributed computing software package known as SeisCorr, which enabled the efficient search of multiple years of continuous data for the occurrence of waveforms similar to those of thousands of template events. We discuss our method of choosing template events and introduce an approach for setting correlation detection thresholds to achieve a given false alarm rate (FAR). With properly chosen templates and corresponding thresholds, SeisCorr found thousands of detections that we were able to validate at each station during our chosen time period, most of which were not in global or regional bulletins created using many stations.

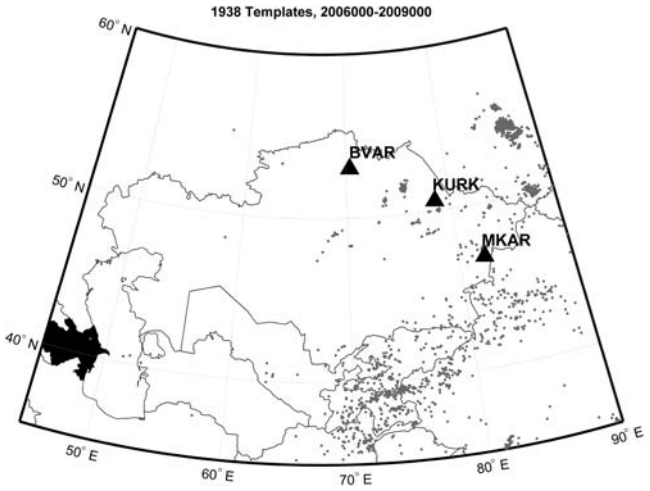

Figure 1. Geographical area of study. LEB events in this region that became templates are shown in gray.

Validation of waveform correlation detections can be challenging. Because waveform correlation can detect smaller events than energy detectors (such as detectors based on a short-time average compared to a long-time average [STA/LTA]), it is not uncommon to detect a waveform with no visually identifiable arrival. Knowing which detections are valid and which are noise detections can therefore be problematic. We discuss and evaluate three methods of confirming detections: bulletin confirmation, multistation validation, and high correlation with a template. Using the latter two methods, we confirmed thousands of detections that are not in event bulletins for the region.

Waveform correlation also enables precise assignment of relative magnitudes (Gibbons and Ringdal, 2006; Schaff and Richards, 2011, 2014), so we were able to assign magnitudes to the events that correlated well with a template. We found ranges of up to three magnitude units within events found by the same template.

In this article, we detail methods for selecting templates (e.g., the choices of phase [first *P* arrival, *Lg*, or both], time window, and frequency band) for setting correlation thresholds and for validating events, and we give a sense of the computational challenges in searching long continuous records for the occurrence of previously recorded waveforms. Our work shows that (1) broad regional monitoring is feasible, (2) correlation thresholds can be set precisely, (3) multistation confirmation of detections is robust, and (4) broad regional monitoring done in this manner yields large numbers of confirmable detections.

Dataset

To investigate the potential utility of waveform correlation for seismic event monitoring at regional and near-teleseismic distances, we focused on a part of central Asia centered on Kazakhstan (Fig. 1) for which there is an openly available regional bulletin of seismicity (principally earthquakes) made available by the Kazakhstan National Data Centre. We processed three years (2006–2008) of continuous waveform data from the International Monitoring System (IMS) stations in the region and used the International Data Centre Late Event Bulletin (LEB) catalog as our source of candidate events from which to seek waveform templates (Table 1). We selected a small set of stations to process based on capability to record signals for LEB events in this region. We found that short-period vertical component array stations MKAR, ZALV, BVAR, and broadband three-component station KURK had the highest percentage of observations of LEB origins from our region, with 95%, 74%, 66%, and 56% of the origins having an arrival (*P* or *Lg*) picked, respectively. We chose not to use ZALV due to instrumentation changes midway through our time period.

Table 1
Dataset Used in This Study

Dates (yyyy/mm/dd)	2006/01/01–2009/01/01
Stations	MKAR, BVAR, KURK
Channels	9 (array), 9 (array), 3 (3C)
Latitude/Longitude box for template acquisition	Latitude: 35°–60°; longitude: 45°–90°

We performed array processing by correlating the incoming data stream on each channel with the recorded template waveform on that channel. Results from each channel were then averaged; we require at least three channels to have contributed to this average (e.g., for a nine element array, only up to six channels can contain data dropouts). The time delay from the incoming signal moving across the array is captured in the individual sensor templates, thus, a high-array correlation value reflects not only similar waveforms at each element, but also appropriate delays across the array. Three-component processing is done similarly; the three components are correlated separately and the results averaged, and all three components must be present and contribute to the average.

Computational Infrastructure

All processing was done using Sandia National Laboratories' SeisCorr software. SeisCorr allows users to build sets of templates and to correlate those templates against continuous data; it also interfaces with our Oracle database to read in data and write results. It was designed to run in parallel on a cluster of commodity computers, meaning that large template sets and data streams can be efficiently processed by adding inexpensive hardware rather than needing to purchase expensive specialized hardware. To achieve optimal efficiency, our correlation algorithm is performed in the frequency domain as described in the Appendix. SeisCorr is written in JAVA.

For each run, which gathers results for one station, SeisCorr performs the following process:

1. Distribute the set of templates across the cluster such that each cluster machine is responsible for a small set of the templates.

2. Divide the continuous data stream into buffers (typically 12 min in duration). Each buffer is sent to all of the machines in the cluster.
   - Send multiple buffers at a time and queue them up on the processing machines to reduce the time to transfer the data across the network.
3. On each machine, correlate the buffers with all templates stored on that machine.
4. Combine the correlation results from each machine.
   - Keep only detections that are above the correlation threshold determined for each template. Find the local correlation maxima (only one detection is allowed in the time duration of the template).
   - When two or more templates make the same detection, associate the detection to the highest scoring template.
5. Write the detections to the database.

Processing performance is dependent on five factors: the number of templates, the number of points in each template (a combination of the duration of the template and sample rate), the number of data-stream points to be processed (a combination of the duration of the data stream and sample rate), the number of processors being used, and the system architecture.

For the typical template library size used in this study (~1000 templates), we found that no additional benefit was gained by using more than 32 processors. At that point, each processor had an optimal number of templates to process; beyond 32 processors, the system was no longer processor bound, and thus additional processors did not add any benefit. Using this number of processors, processing the three-year period, for 21 channels, took 2.5 days total.

Template Selection and Screening

Template selection is a critical aspect of any waveform-correlation-based system. Our experience has shown that, if the goal is operational use, the need for care in template selection can hardly be overstated. A well-running waveform correlation detector needs clean, clear template waveforms, without noise artifacts (these can trigger detections on noise segments). Forming templates requires finding waveform candidates from historical events, then deciding which phase(s) and time segments (window lengths) of the waveforms to use, and finally screening the waveforms to eliminate noisy or otherwise unsuitable templates. To find candidate template events for a given station, we queried the LEB arrival and origin tables for all arrivals associated to events in our latitude/longitude box that occurred during our time period.

Figure 2 shows representative waveforms from our BVAR template library as a function of distance between station and origin location. This type of distance versus time plot allows us to visually inspect the collection of waveforms for a station, note the most prominent arrivals, and select reasonable window placements to use for extracting our templates.

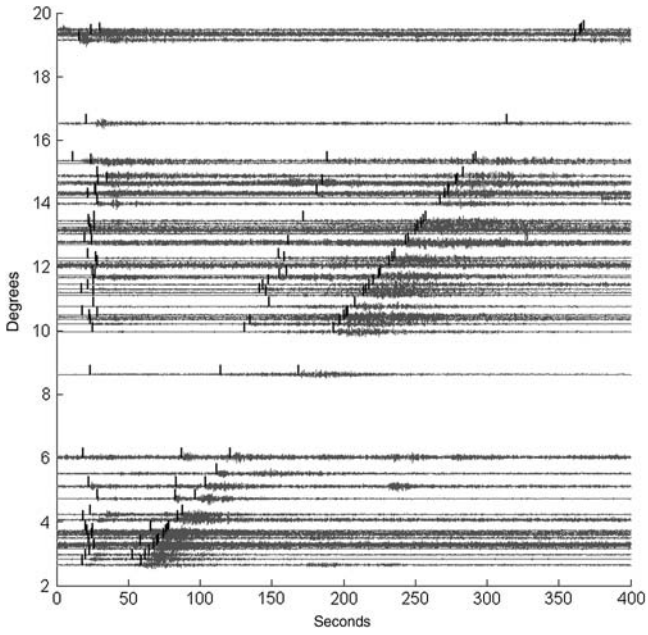

Figure 2. Profile plot of BVAR templates. Only a subset was plotted to improve visibility (one waveform for all events with station-to-event distances within 5 km) and waveforms were time shifted to align on the theoretical first *P*. LEB arrivals are marked by a black bar; *P* arrivals occur at about 20 s; *S* arrivals are the middle group; *Lg* arrivals are the last arrivals, with higher amplitude.

In future work, we anticipate customizing the window length to each template so that templates from events farther from the station will have a longer window length, etc. However, at this stage in our research, we opted to use the same window length across the template library for each station.

Choosing the optimal window length depends on the choice of phase or phases and vice versa. We compared three options for the choice of phase(s): first *P* (*Pn*, *P*), *Lg*, and first *P* through *Lg*. For first *P*, we used 40 s windows (starting 5 s before the first *P*); for *Lg*, we used 25 s windows (starting 5 s before the *Lg* arrival); and for first *P* through *Lg*, we started 5 s before first *P* and went until 25 s after *Lg*. To guide us in the choice of window length and phases, we found 99 events with both good first *P* and good *Lg* arrivals. Next, we made templates from these events using the three arrival window options. We set the correlation threshold for each template right at the cusp of consistently getting good, visually verifiable matches. We processed one year of data with each of the three windowing options and compared results. The *Lg* templates found significantly more detections than the first *P* templates: 750 compared to 322. Most of the detections based on the first *P* template (293) were also detected using the *Lg* templates. We suspect this is due to the higher signal-to-noise ratio (SNR) of *Lg* (evident in Fig. 2) and greater duration and bandwidth. *Lg* is commonly the strongest regional seismic phase from shallow continental events (Press and Ewing, 1952). Our templates of first *P* through *Lg* performed as well as using just *Lg*; however, the longer window is more computationally expensive. Therefore,

Table 2
Summary of Templates Obtained at Each Station

Station	Phase	Filter (Hz)*	LEB Orids in This Region between 2006/01/01 and 2009/01/01†	Initial Template Candidates (had LEB *Lg* Arrivals)	Candidates that Passed STA/LTA Screening‡	Templates after Clustering§
MKAR	*Lg*	0.5–5	4886	2165	1023	811
BVAR	*Lg*	0.5–5	4886	644	586	543
KURK	*Lg*	0.5–5	4886	3234	1881	1515

*Third-order Butterworth, applied forward and backward.
†Origin identifier.
‡Short-time average (STA) = 1 s, long-time average (LTA) = 30 s, gap = 0.5 s overlap; STA/LTA threshold = 3.
§Agglomerative clustering; threshold to select families was 0.7.

we decided to use *Lg* templates for this study. We do note, however, that *Lg* is not always strong and can be blocked by changes in crustal structure. Events in the Caspian Sea region tended not to produce *Lg* arrivals at our stations. In a more highly developed approach, one could form a template library that mixed *Lg* and *P* arrivals, but for our initial investigation, *Lg* arrivals were satisfactory.

Thus, for this study, to form template libraries at each station, we first queried the LEB bulletin for *Lg* arrivals at each station from events with origins in our time period and our Kazakhstan region. Our candidate waveforms were then filtered to enhance *Lg* (sixth-order band-pass filter keeping 0.5–5.0 Hz) and windowed to capture 25 s starting 5 s before the *Lg* arrival. Next we screened using a two-stage process to find clear, clean, representative waveforms. First, we used STA/LTA to verify that there was an observable arrival in our waveform segment (STA was 1 s, LTA was 30 s, gap was 0.5 s, STA/LTA threshold was 3). Second, we applied agglomerative hierarchical clustering (dendrograms) to the waveforms to identify families of similar events (Merchant, 2007). Events that clustered with a correlation value above 0.7 were considered a family. We clustered to avoid having multiple templates representing the same location and source type (e.g., the same mining pit), so we chose one representative template for each family. We kept the waveform with the highest average correlation with other events in the family as the representative template for that family. If there were only two events in a family, we kept the event with the higher STA/LTA value. The template candidates that passed through this screening process became our template library.

At MKAR and KURK, less than half of the waveforms with LEB *Lg* arrivals passed our STA/LTA screening process, suggesting that analysts routinely include *Lg* phases with poor SNR. This seems quite reasonable because *Lg* arrival time is typically not used to determine location and hence need not have an impulsive arrival. Clustering the templates that passed STA/LTA screening further shrank our template library. To understand how our processing results compare between stations, it is important to note that the template libraries are generated independently for each station. As a result, they do not necessarily contain the same events. We could have required template selection to use the same events for all stations, but this would have limited the number of events that could have been used as templates because not all events have an *Lg* arrival that passed STA/LTA screening at each station, and it would have decreased the overall detection capability at each station because the best correlating *Lg* template for a cluster might not come from the same event at each station. A summary of all our template selection parameters and decisions is shown in Table 2. Figure 3 shows the intersection of the template libraries at our three stations. Figure 4 shows the effect of the template-winnowing process at MKAR. The complete set of all event locations for templates used in this research paper, from all stations, is shown in Figure 1.

Threshold Selection

The final, critical step to characterize a template library for use by our waveform correlation detector is to determine a correlation value to serve as a detection threshold for each template. For an operational system, the preferred method is to set a desired FAR and derive the threshold from that. In some situations, we may want to guarantee that all detections are valid and high quality; in other situations we want to find all candidate similar events, including those buried in noise, and are happy to accept some false detections in order not to miss anything of potential importance. Wychecki-Vergara *et al.* (2001) developed a statistically rigorous method to set correlation thresholds for templates to achieve a specified FAR, but they assume that the noise is Gaussian, which is not generally valid for our datasets, and we found that the resulting thresholds lead to FARs that are significantly higher than expected.

We have found better success expanding on a technique based on processing time-reversed templates developed by D. Schaff (unpublished manuscript, 2010). Intuitively, when we think about the distribution of correlation values generated by a waveform correlation detector operating on continuous data, we can think about the distribution as the sum of two distributions: the contribution from correlations with noise windows, which is centered about a correlation value of zero, and the contribution from correlations with similar waveforms that is produced by the events we are trying to detect. If we knew the distribution of the contribution from noise windows alone, this would allow us to estimate how many false detections to expect at a given correlation threshold. The challenge, then, is to generate this distribution isolated

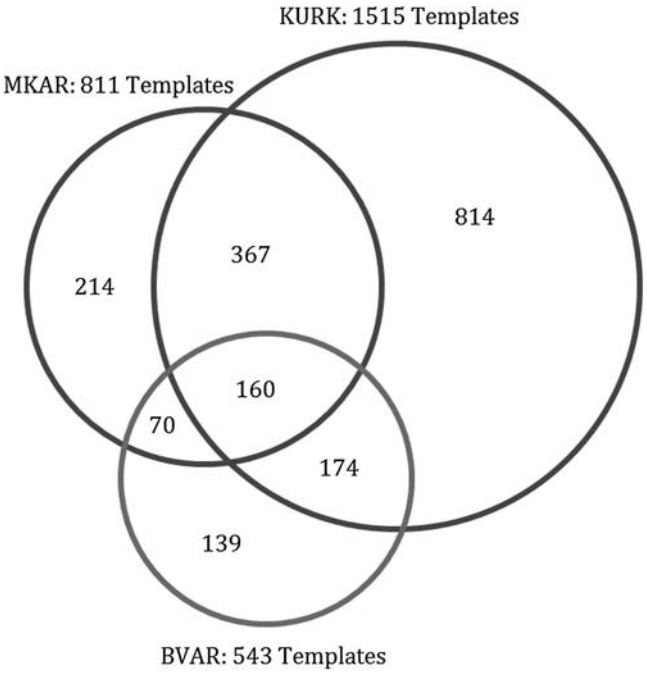

Figure 3. Venn diagram of template libraries at each station. The intersection of template libraries at our three stations is shown; 160 events were selected as templates at all three stations, 214 were unique to MKAR, 367 were common to MKAR and KURK but not BVAR, etc. There were 1938 unique seismic events that generated templates.

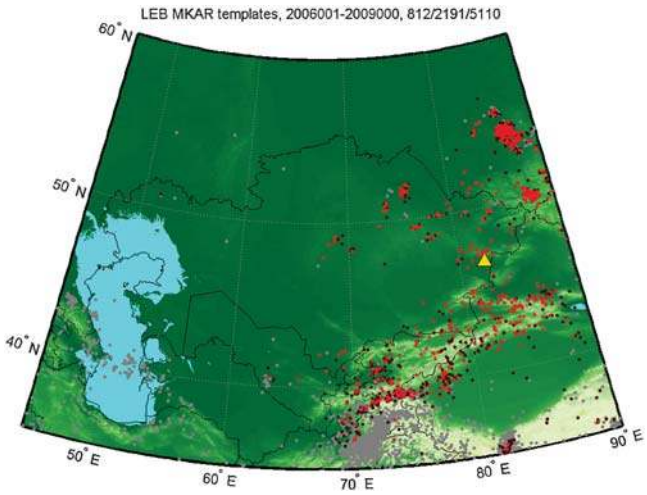

Figure 4. Our template screening process for MKAR. Events in the LEB catalog that did not have *Lg* arrivals at MKAR are shown in gray; events with an *Lg* arrival at MKAR that did not pass screening are shown in black; events in our final template library after short-time average/long-time average screening and clustering are shown in red.

from the signal correlation contribution. A theoretically simple approach would be to remove all the data segments with matching signals from the continuous data, then perform the correlation, plot the distribution, and set a threshold to achieve the desired FAR. The problem, of course, is that we do not know how many such signals there are or when they occur: that is the point of running the detector in the first place.

Correlating time-reversed templates with continuous data provides a simple way to generate the distribution from noise with no need to delete any time windows. Reversing the template ensures the template has the same time-bandwidth product as the template proper and should yield a very similar distribution of correlation values in noise windows. At the same time, by using a reversed template, we should ensure that no time-forward matching signals will ever be correlated with the template, hence there is no second distribution of the reversed template with matched signals. We confirmed empirically that time-reversed templates used in normal operation produced correlation outputs with the same distribution as time-forward templates during times of background noise. This method makes no assumption about the distribution of noise or signal data as being Gaussian or any other standard distribution. The only assumption that we make in this study is that the noise is fairly static: we did not vary the detection threshold within the three-year period. However, thresholds could be dynamically and empirically adjusted for shorter time periods (during strong aftershock sequences, for example), using an extension of this method.

Typical examples of the time-reversed and time-forward template distributions for a single MKAR template are plotted in Figure 5. The distributions from the forward and reversed templates are overlaid and indistinguishable at the lower correlation values corresponding to the noise window correlations. The differences are all in the higher correlation value ranges, which are entirely lacking from the reversed template results; only the forward templates triggered high correlation values. This result is consistent for all of our templates, though there is considerable variation in the shape of the distribution generated by each template.

Figure 6 shows the number of detections by SeisCorr at array station MKAR over the three-year period as we vary the correlation threshold for this pair of time-forward and time-reversed templates. Looking at detections is slightly different than looking at the distribution of correlation values: a similar event waveform passing through SeisCorr generally yields many high correlation scores as peaks and valleys align as it passes by, but SeisCorr will reduce these to one detection corresponding to the single window that best matches the signal. At high-correlation threshold values, only the time-forward template triggers detections. As the correlation threshold is lowered, the number of detections rises slowly. Around correlation threshold 0.27, the time-reversed template starts producing detections and the distribution of time-forward template detections changes slope and then starts climbing rapidly, in parallel with the time-reverse template detections numbers. The obvious interpretation is that at correlation threshold value 0.27, SeisCorr starts triggering on noise. Time-reversed templates give us a method to (1) find the point at which each individual time-forward template will start triggering on noise and (2) allow us to set our correlation thresholds to achieve a desired FAR.

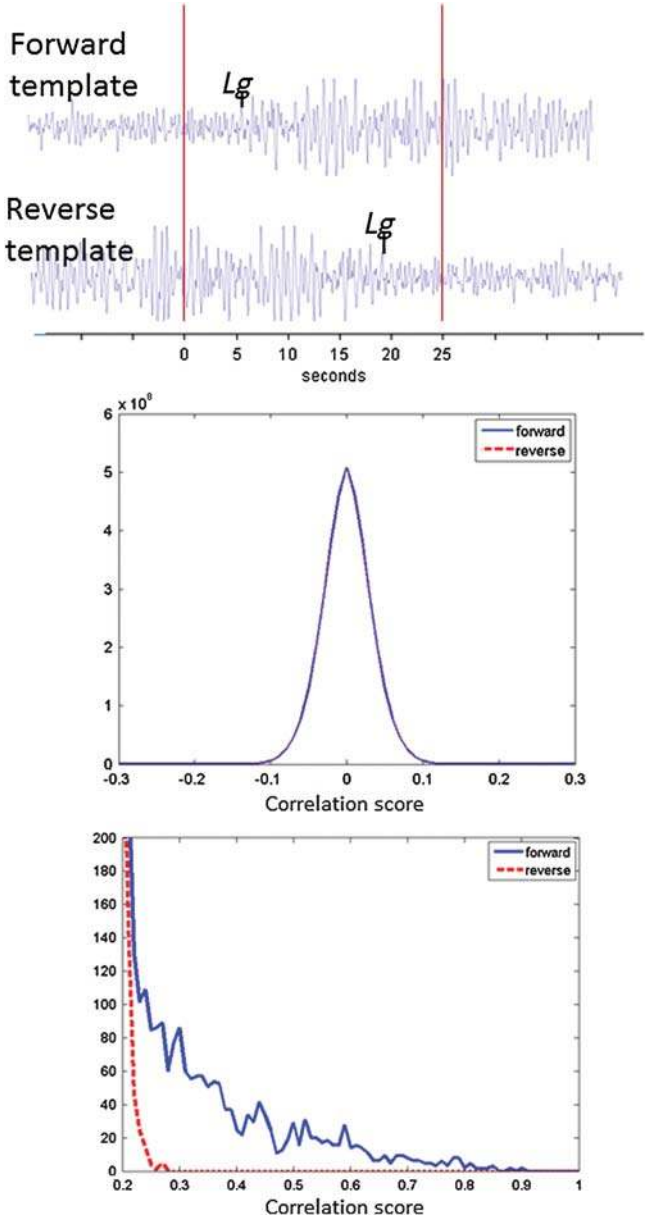

Figure 5. The top figure shows the waveforms for the time-forward and time-reversed template at MKAR from origin identifier (orid) 3548148. The middle figure shows that the distributions of correlation values are indistinguishable at this zoom level and essentially identical. The bottom figure shows that only the forward template ever correlates at values above 0.27.

To obtain template thresholds for our system, therefore, we take each template, time-reverse it, run it on the three years of continuous data that we are processing to look for similar events, and generate the distribution by forming a histogram of all correlation values. A correlation detection threshold can then be selected based on the desired FAR. In this manner, all templates can be assigned a correlation threshold that is consistent in FAR, improving on the practice of using the same correlation threshold for all templates, and allowing consistency across templates of different window lengths or filtering parameters. For this study, we calculated

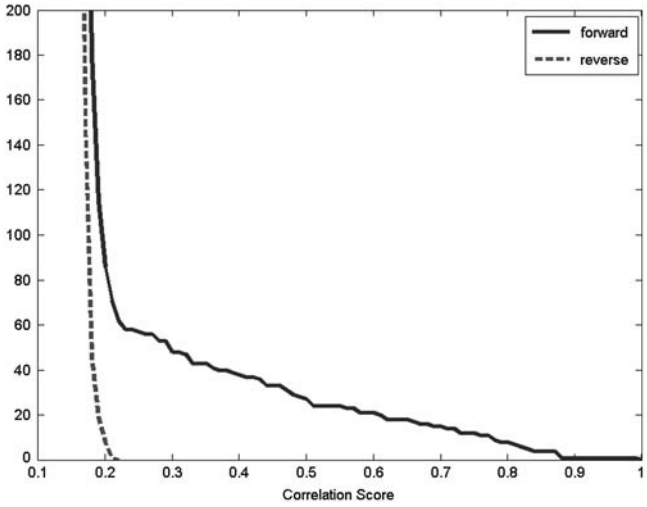

Figure 6. Number of detections triggered in SeisCorr by the forward and reversed template as correlation detection threshold is varied. Template from orid 3548148 at MKAR.

correlation detection thresholds for each template that resulted in three false alarms during the three-year time period, that is, a FAR of one false alarm/year. The importance of setting thresholds on a template by template basis is demonstrated by the variability of thresholds for the templates used at MKAR: template threshold correlation values for a FAR of one per year ranged from 0.16 to 0.63. At the high end are templates with very low time-bandwidth products. By calculating thresholds in the manner we have described, one can set a consistent FAR for all templates using the actual noise conditions at each station.

One additional subtlety to the topic of threshold selection is that incoming events do not have a binary declaration of similarity, or lack thereof, with the template event. Our original description of the distributions (one consisting of noise and one from similar events) glossed over the fact that there are degrees of similarity. One topic of concern in this research area is always that an unrelated event, especially one from a similar distance, could have a higher correlation value with the template than noise does but would be considered a false detection for operational purposes. This is less of an issue in our dataset where we used fairly broadband *Lg* but could be a problem if templates were fairly narrowband and relied heavily on capturing the *P*-to-*S* separation. Because of how this method sets the threshold, it usually results in high thresholds for narrowband templates prone to this problem, which helps minimize such undesired detections. Looking at the number of detections generated at each threshold level (Fig. 6) and the slope of that curve can aid selection of a proper threshold for a given operational use if detections of this type are a problem. Lastly, the validation method described in the next section also mitigates this, assuming that stations are decently far from one another and therefore unlikely to have the same template detected on the same undesired event. This topic will surface again when we describe our final threshold selection.

Validation of Detection Using Multistation Analysis

Although waveform correlation is a robust detection method that can be set to have a low FAR, continuous data involve so many correlation calculations that even when running a waveform correlator with a low FAR, the number of false detections can be significant and even outpace true seismicity rates. The number of false alarms obtained when processing continuous data scales linearly with the number of years, sample rate (assuming one correlation score is calculated for each sample start time), number of templates, and number of stations. For example, at MKAR over three years with 811 templates, we would anticipate 2433 false detections even if each template generated only one false alarm per year. (For comparison, the LEB catalog listed 5110 events seen at MKAR from our region in that time period.) As operational scenarios would involve running a correlator against continuous data, additional validation techniques would be beneficial.

Fortunately, the probability of more than one station detecting a false event at the same time using waveform correlation is very small. To perform multistation validation, we calculated origin times for the detections at each station, assuming that the travel times for the detections were the same as the travel time for the template. Because one of the basic premises of waveform correlation is that template events and detected events are nearly collocated, this is a reasonable assumption. If a detection from one station had a calculated origin time within 4 s of the calculated origin time of a detection at another station, we assumed the detections were from the same event and called it a validated detection.

The probability of a false detection within a 4 s window at a station with one template with the correlation threshold set so that it has 1 FA/year is

$$p1 = \frac{1\ \text{FA}}{\text{year}} \times \frac{\text{year}}{365\ \text{days}} \times \frac{\text{day}}{24\ \text{hours}} \times \frac{\text{hour}}{3600\ \text{s}} \times \frac{4\ \text{s}}{\text{window}} = 1.27 \times 10^{-7}.$$

If we assume the rate of false detection is independent at each station (probably not totally true, but a decent assumption if our stations are reasonably far apart), the probability of two stations making a false detection within that 4 s window from the same template is

$$p2 = p1^2 = 1.6 \times 10^{-14},$$

and the probability of three stations making a false detection in that time span is

$$p3 = p1^3 = 2.0 \times 10^{-21}.$$

Thus, the probability of a false detection plummets as more stations are required to confirm.

With more templates, the probability that one of them makes a false detection goes up. If there are 1000 templates at each station, the probability of a false detection in a 4 s window is

$$P1\_1000 = 1000 \times p1 = 1.7 \times 10^{-4},$$

$$P2\_1000 = p1\_1000^2 = 1.61 \times 10^{-8},$$

$$P3\_1000 = p1\_1000^3 = 2.04 \times 10^{-12},$$

which is higher but still low as it translates to one false two-station confirmation about every 8 years and one false three-station confirmation about every 62,000 years.

These numbers give us a sense of the reduction in false alarms expected by the use of multistation analysis. By restricting the detecting templates to be within a certain distance of one another, we would operate between these two example cases: a station might have 1000 templates to cover a broad geographical region, but generally only a much smaller number of templates would be within a given distance of the template that found the detection. Although correlation detection at a single station can produce unwieldy numbers of false alarms compared to the true seismicity rate, combining correlation detection with multistation validation gives us a very high degree of confidence in our validated detections.

In this dataset, we can study the number of false alarms directly by looking at the number of detections obtained when we correlated the time-reversed templates with three years of data and then seeing how many of these are confirmed using multistation analysis. Our results show that although our large libraries of time-reversed templates generate a number of false detections at each station during the three-year period, as was expected from thresholds being set to allow 1 FA/year, the false detections rarely occur at the same time (Table 3). The detections at BVAR and KURK that had the same calculated origin time were from templates with LEB locations 1883 km apart, so we would have had no multistation confirmations if we required the detecting templates to be reasonably collocated. We thus believe multistation validation is a robust method to validate waveform correlation event detections.

The reader might note, upon close study of Table 3, that the number of detections obtained by single stations was actually lower than anticipated, as we expected three detections per template during the three-year period from our 1 FA/year; this is due to our decision to round detection thresholds up to the hundredth place (e.g., 0.374 would be rounded to 0.38), as our goal was to ensure the desired FAR would not be exceeded.

Threshold Selection for Reported Results

For this study, our initial aim was not to generate as many detections as possible, but to find families of very high-quality detections. Thus, we chose to add a small empirically chosen offset of 0.05 to all our threshold values. Adding the small offset reduces the geographic spread of events matched with a template and ensured that detections are recognizable as similar by an analyst, which fits with our goal to evaluate waveform

Table 3
Detections Obtained during the Three-Year Period from Time-Reversed Templates

Station	Detections by Time-Reversed Templates	Detections Per Template	Two-Station Detections	Three-Station Detections
MKAR (811 templates)	1098	1.35	0	0
BVAR (543 templates)	422	0.77	1	0
KURK (1515 templates)	3408	2.25	1	0

Table 4
Detections Obtained during the Three-Year Period at Each Station

Station	Number of Detections	Detections Confirmed in the LEB	Detections Confirmed in the KNDC Catalog	Detections Confirmed in Either Catalog	Detections Confirmed by Two Stations (1481 were Seen by all Three Stations)	Detections Confirmed by High Correlation*	Confirmed in Some Manner
MKAR (array)	7426 (from 506 of 811 templates)	526	553	927	4740 (64%) (4226 not in LEB)	1309 (183 not in LEB or confirmed by another station)	5136 (69%) (4610 not in LEB)
BVAR (array)	3096 (from 193 of 543 templates)	154	67	180	2199 (71%) (2054 not in LEB)	383 (17 not in LEB or confirmed by another station)	2307 (75%) (2153 not in LEB)
KURK (3C)	8526 (from 837 of 1515 templates)	693	705	1179	4895 (57%) (4327 not in LEB)	2031 (376 not in LEB or confirmed by another station)	5648 (66%) (4955 not in LEB)

*Correlation value > 0.7.

correlation event detection for nuclear explosion monitoring purposes. For example, for the template used in Figure 6, our adjusted correlation threshold was $0.27 + 0.05 = 0.32$.

At these thresholds, the time-reversed templates yielded no detections, even at the single stations. We therefore suspect the single station results presented in this article are quite robust in their own right.

Results and Validation of Results

With template libraries built and thresholds for each set as described above, SeisCorr detected thousands of similar events at each station (Table 4). The number of detections at each station is listed along with the number of templates that found those detections. (Not all templates ended up finding similar events, thus we list both the number of templates that made detections and the total number of templates that passed our screening criteria and were in the template library. For clarity, we note that when processing the data stream, the nonperforming templates of course detected themselves with a correlation score of 1; such self-detections are not included in our detection counts, however.) Based on the concentration of some detection families in only certain times of the day, we concluded that many of the families (including all of the largest ones) correspond to mining events. Figure 7 shows a set of waveforms for a typical mine family, and Figure 8 shows the corresponding time-of-day histogram. It is also worth noting that our best performing templates were often those with fairly low calculated correlation thresholds; these templates had more complex signals and higher time-bandwidth products. Poorer quality templates often are more narrowband and had high calculated correlation thresholds. Setting a constant threshold would have missed many of our best signals in favor of lots of false detections from the poorest templates.

To evaluate the validity of the detections, we used three methods. In our first method, we searched for a corresponding origin in a catalog at our calculated origin time (within 4 s). We applied this method using both the LEB catalog and the regional catalog from the Kazakhstan National Data Center (KNDC). The KNDC catalog has many more events than the LEB in our processing time period, and we hoped that it would corroborate many of our detections. However, the KNDC catalog is primarily intended as a list of earthquakes, for purposes of studying seismic hazard and tectonics, so it screens out most mining explosions, which seem to be the most common type of event that we detected in the continuous archive. The highest number of confirmed events for a station in the KNDC catalog is 705 at KURK, but that is still only 8% of the total events detected at KURK. This is only slightly better than the 693 events confirmed via the LEB. The events confirmed by the different catalogs had a fairly small overlap, however, so 13.8% of events at KURK could be confirmed by one of the catalogs.

A second method of validation, multistation validation as described above, does not rely on availability of a catalog. The method compares detections between the stations: if two stations made a detection with calculated origin times within 4 s, we consider the detection to be validated. We found this method worked quite well in practice: the number of validated detections was quite stable as we varied the allowed time difference and the detecting templates' allowed distance, and all

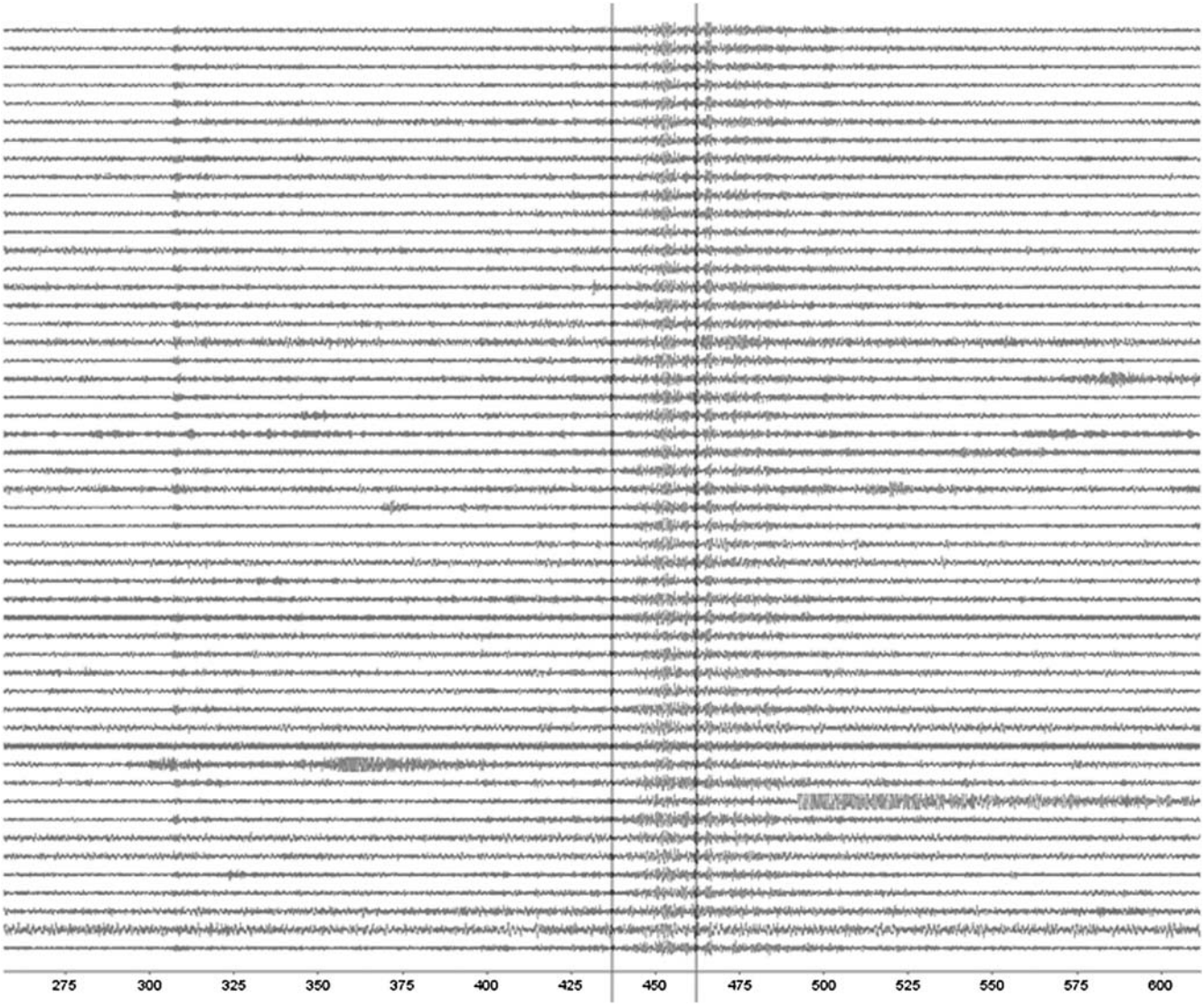

Figure 7. Mining family found by the template from orid 3548148 as seen by one channel of MKAR. The correlation window used was 25 s, starting around the *Lg* arrival at 435 s and indicated with the vertical black lines. The *P* arrivals line up, as well as the *Lg*.

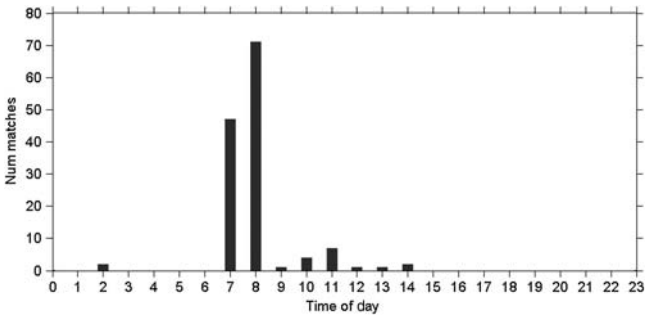

Figure 8. Histogram of the time of day (local time) at which detections were found for the family in Figure 7. All but two detections were during business hours, and most detections occurred at 7 a.m. or 8 a.m. This behavior indicates that these events are human-induced, with mining activity representing the most likely explanation.

of the random samples we selected from our overall results were confirmed by an analyst.

A third method, which also does not require the use of a catalog, is to count detections with extremely high correlation values (>0.7) as confirmed, even if they are seen at only one station. This can be an effective method of validating smaller events that were only seen at one station. However, because the vast majority of detections at such a high correlation value were also confirmed by one or both of the other two methods, this added very few additional confirmed events. This suggests this type of validation is reliable when dealing with very sparse networks, where events may have only been seen at a single station.

Discussion

Most (66%–75%) of the thousands of detections at each station can be confirmed by one or more of our three vali-

Table 5
Unique Detections Obtained during the Three-Year Period from the Three-Station Network

Number of Unique Detections	Detections Confirmed by the LEB Bulletin	Detections Confirmed by the KNDC Bulletin	Detections Confirmed in Either Bulletin	Detections Confirmed by Two or Three Stations	Detections Confirmed by Three Stations	Detections Confirmed by Single Station High Correlation*	Confirmed in Some Manner
12506	976 (15%)	973 (15%)	1637 (25%)	5058 (77%)	1481 (23%)	2777 (42%) 494 (7.5%) confirmed solely by high correlation	6563 (4926 not in the LEB or KNDC bulletin)

The number of events confirmed via each method is shown, along with the percent of confirmed events that were validated in that manner.
*Correlation value > 0.7.

dation methods: bulletin comparison, multistation validation, or an extremely high single-station correlation value. Multistation validation confirmed the most events: 57%–71% of the detections at each station were confirmed by at least one other station. Looking at the three stations as a network (Table 5), we confirmed 6563 unique events. Again, multistation validation confirmed the majority of these events (77%). Almost a thousand (976) of our detections were of events that were in the LEB, which means waveform correlation is detecting events that the monitoring community finds to be of interest. In addition, we find thousands of detections not already included in either the LEB or KNDC catalogs (though we note that many of these events may have been detected by the KNDC and then screened out as mining events). Adding the 4926 confirmed events that were not in the LEB or KNDC bulletins would almost double the LEB catalog size, as the LEB contained 5110 origins in our region and time period.

The magnitude distribution of these new confirmed detections is of significant interest. Relative magnitudes of detections in comparison to the template were calculated during the correlation process following the method of Gibbons and Ringdal (2006).

Relative magnitude is estimated as $\log \alpha$:

$$\alpha = \frac{x.y}{x.x},$$

in which x is the template waveform and y is the incoming data waveform window that contains the similar event.

Figure 9 shows the distribution of calculated magnitudes of detections at MKAR. We assign an absolute magnitude for each detected event by adding the calculated relative magnitude (based on *Lg*) to the detecting template's LEB m_b magnitude. We note that the magnitude distribution of confirmed detections was shifted toward larger values than the distribution for detections as a whole, which is expected because the smaller events are less likely to be confirmed by multiple stations or by catalogs. Moreover, the largest detections could frequently be linked to events in the LEB. This relative magnitude calculation slightly underestimates the magnitude for detections that are not perfectly correlated with the template (Schaff and Richards, 2014); this could be why our calculations are slightly lower than the m_b values listed in the LEB catalog for the same events. Nevertheless, it is significant that we are detecting so many events that are a magnitude unit smaller than the LEB events, and this confirms observations of Gibbons and Ringdal (2006) and Schaff (2010) that waveform correlation can detect events up to one magnitude smaller than conventional methods.

Figure 10 shows the histograms of calculated m_b magnitudes for all detections in the mining family shown in Figure 7. In this typical example, the template waveform had an m_b of 3.1, and the family of detected events spanned 1.6 magnitude units. As expected, most of the detected events were smaller than the template. The largest events from a mine are the ones most likely to be in the LEB catalog, the source of our templates. Our clustering process is also likely to choose the largest event in each family as its representative.

As we contemplate expanding from broad regional monitoring to global monitoring, it is relevant to discuss whether our results from this region of the world are typical. A key part of our success here is due to the large number of large mining events; based on time of day histograms, our largest families all appeared to be from mines. All stations had several families of over 100 detections; there were 137 families with more than 40 detections. Not all of our detections are from mines, however, and we had 971 families of less than 10 events. Nevertheless, regions with less mining activity and/or with smaller mining explosions may find less benefit, and we need to study other regions to learn if our results here are typical and can be considered as representative of global behavior.

In our study, we detected 976 events that were in the LEB bulletin, about 20% of the bulletin. The benefits of detecting these events using waveform correlation rather than traditional methods include having immediate access to estimated origin time, location, and phase arrival times. Robust analysis of the benefits of using correlation to detect events for the bulletin would be better done with a study that used historical templates and added templates in chronological order; however, our results suggest a significant portion of the events in the LEB catalog may be detectable via correlation. Adding in templates drawn from a longer time period (we used only three years in the study period in this preliminary study), using a subspace detector (Harris, 1991; Harris and Dodge, 2011), and increasing station density should all have the effect of further increasing the number of detected events. In future work, we intend to study the portion of published

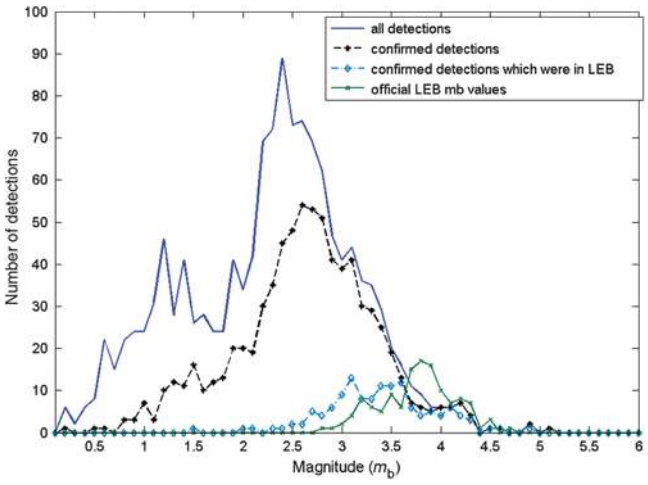

Figure 9. Calculated magnitudes of detections. Number of events per 0.01 interval in magnitude.

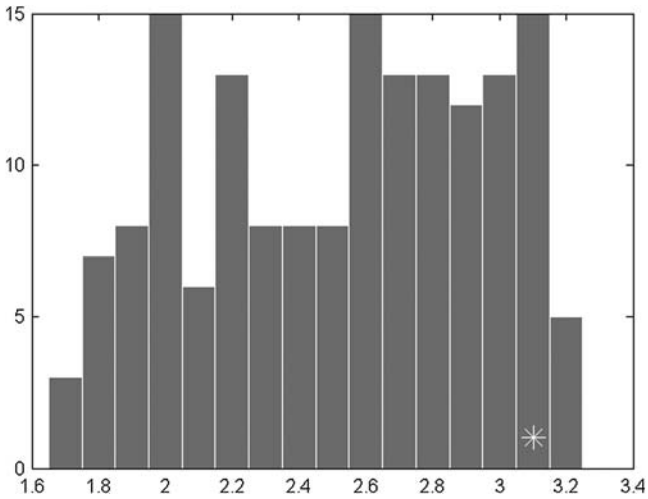

Figure 10. Histogram of calculated magnitudes of events in a mining family detected at MKAR. The asterisk indicates the magnitude of the template for this family (3.1).

bulletins which could be detected using waveform correlation. Possible refinements to our template generation process could improve the detection capability as well. Possibilities include improving our screening criteria and using variable window sizes so that each template has only the highest SNR portion of the recorded waveform.

For the monitoring scenario explored in this study, we wanted detections to be of very high quality. Running time-reverse templates at our final threshold values triggered no detections during the entire three-year period; this is a good proxy for the number of false alarms expected from the time-forward templates. This, and our high percentage of confirmed detections at each station, gives us confidence in our detections and indicates that our time-reversed template threshold selection method works well. An event that is not validated can still be a true detection of a similar event; it may just have a magnitude too low for it to be seen at another station or included in a bulletin. Another method of validation which might help screen the as-yet unvalidated events is doing postdetection frequency–wavenumber (f-k) screening on the correlation traces (Gibbons and Ringdal, 2006; Gibbons *et al.*, 2012). Additionally, although we were pleased with the performance of our correlation detection thresholds, it is worth noting that in situations of variable noise, an adaptive approach of scaling the cross-correlation coefficient values such as that used by Schaff (2008, 2010) could be integrated and might be beneficial. Instead of setting our correlation thresholds based on the whole three years of data, we could set thresholds based on the recent past. Accordingly, an adaptive reverse-template approach is a distinct possibility in future versions of our software.

Conclusion

Using our SeisCorr software, we performed broad regional monitoring of central Asia. In two and half days we processed three years of continuous waveform data (21 channels) at three stations against templates from 1938 unique events. We detected thousands of similar waveforms at each station (3096 [BVAR], 7426 [MKAR], and 8526 [KURK]); most (51%–71%) of these detections were detected at more than one station or were otherwise validated (66%–75%). In all, we detected 5058 unique events that were confirmed by another station and 1481 seen at all three. Our results support two preliminary conclusions: (1) waveform correlation can detect many events in current bulletins as well as thousands of additional similar events that are currently going undetected, and (2) we can trust the detections we are obtaining.

Almost a thousand events in the LEB were detected in the continuous waveform data by SeisCorr at each station, by templates drawn from the same time period. A thousand events are about 20% of the catalog. With templates drawn from a longer time interval and/or the use of a subspace detector, we suspect this percentage could be higher. Our detection of thousands of confirmable detections not currently in existing bulletins points toward the power of waveform correlation; as bulletin-reporting requirements move toward ever-smaller events, waveform correlation provides a method of detecting a significant number of those events in an organized and efficient manner. In addition to having a detection time, an event detected using waveform correlation also has an estimated origin location, an estimated origin time, and estimated arrival times for all phases at each station that detected the template event. Preliminary work at Sandia integrating our SeisCorr software with an associator indicates that providing waveform correlation event detection results to an associator can improve event building significantly. Another benefit of using waveform correlation to detect events is the ability to used results to do relative event location (Waldhauser and Schaff, 2008). Moreover, in an operational environment, if families of similar events can be grouped and given context before being handed to an analyst for processing, rather than having an analyst processing each event individually, time savings could be significant.

For eventual application to global monitoring, we would need to carefully select template thresholds. We presented a method for determining thresholds on a template-by-template basis with a desired FAR. Lower thresholds on high-quality templates and higher thresholds on low-quality templates allow the detection of a plethora of good detections from the best templates while minimizing false detections from the poorer templates. In this study, we were able to validate 66%–75% of our detections, which gives us a lower bound on the number of valid detections because unvalidated events are not necessarily noise or incorrect detections.

The results of our study using our SeisCorr software to process data from IMS stations in central Asia suggest that doing broad regional monitoring using historical and real-time-generated templates is feasible and useful. Robust detections of new events and the additional information one can derive from these detections suggest that waveform correlation can improve upon traditional monitoring schemes by allowing detection of smaller events, fewer errors in association, and faster event processing.

Data and Resources

Waveform data for International Monitoring System stations MKAR and BVAR were obtained from the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) via the United States National Data Center. Data are available to all member states via their national data centers. Our data were downloaded on May 2012.

Waveform data for KURK can be obtained from the Incorporated Research Institutions for Seismology Data Management Center at www.iris.edu (last accessed July 2013).

The Late Event Bulletin was obtained from the International Data Centre of the CTBTO and is available to all member states. Our data was last downloaded May 2012.

The regional seismicity bulletin that we used is a product of the Kazakhstan National Data Centre and can be downloaded from http://www.kndc.kz/index.php/en/seismic-bulletins/interactive-bulletin (last accessed September 2014).

Acknowledgments

Some of this work was performed under the auspices of the U.S. Department of Energy by Sandia National Laboratory under Award Number DE-AC04-94AL85000. It was supported also by the Defense Threat Reduction Agency under Award Number HDTA1-11-1-0027 to Columbia University. This article is Lamont–Doherty Earth Observatory Contribution Number 7815. We are grateful for helpful comments from Doug Dodge and an anonymous reviewer.

References

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165,** no. 1, 149–166, doi: 10.1111/j.1365-246X.2006.02865.x.

Gibbons, S. J., and F. Ringdal (2012). Seismic monitoring of the North Korea nuclear test site using a multichannel correlation detector, *IEEE Trans. Geosci. Remote Sens.* **50,** no. 5, 1897–1909.

Gibbons, S. J., F. Ringdal, and T. Kværna (2012). Ratio-to-moving-average seismograms: A strategy for improving correlation detector performance, *Geophys. J. Int.* **190,** no. 1, 511–521, doi: 10.1111/j.1365-246x.2012.05492.x.

Harris, D. B. (1991). A waveform correlation method for identifying quarry explosions, *Bull. Seismol. Soc. Am.* **81,** no. 6, 2395–2418.

Harris, D. B., and D. A. Dodge (2011). An autonomous system for grouping events in a developing aftershock sequence, *Bull. Seismol. Soc. Am.* **101,** no. 2, 763–774, doi: 10.1785/0120100103.

Harris, D. B., and T. Paik (2006). *Subspace Detectors: Efficient Implementation*, Department of Energy, United States, doi: 10.2172/898451.

Hauksson, E., and P. Shearer (2005). Southern California hypocenter relocation with waveform cross-correlation, Part 1. Results using the double difference method, *Bull. Seismol. Soc. Am.* **95,** 896–903, doi: 10.1785/0120040167.

Merchant, B. J. (2007). The GNEMRE Dendro Tool, *Report Number SAND2007-6439*, http://www.osti.gov/scitech/biblio/926809 (last accessed March 2014).

Oppenheim, A. V., and R. W. Schafer (1975). *Digital Signal Processing*, Prentice Hall, Inc., Englewood Cliffs, New Jersey.

Press, F., and M. Ewing (1952). Two slow surface waves across North America, *Bull. Seismol. Soc. Am.* **42,** 219–228.

Schaff, D. P. (2008). Semiempirical statistics of correlation-detector performance, *Bull. Seismol. Soc. Am.* **98,** no. 3, 1495–1507, doi: 10.1785/0120060263.

Schaff, D. P. (2009). Broad-scale applicability of correlation detectors to China seismicity, *Geophys. Res. Lett.* **36,** no. 11, 10.1029/2009GL038179.

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,** no. 2, 829–846, doi: 10.1111/j.1365-246X.2009.04446.x.

Schaff, D. P., and P. G. Richards (2011). On finding and using repeating seismic events in and near China, *J. Geophys. Res.* **116,** no. B3, doi: 10.1029/2010JB007895.

Schaff, D. P., and P. G. Richards (2014). Improvements in magnitude precision, using the statistics of relative amplitudes measured by cross correlation, *Geophys. J. Int.* **197,** no. 1, 335–350, doi: 10.1093/gji/ggt433.

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. Seismol. Soc. Am.* **100,** no. 6, 3224–3238, doi: 10.1785/0120100042.

Schaff, D. P., W. Y. Kim, and P. G. Richards (2012). Seismological constraints on proposed low-yield nuclear testing in particular regions and time periods in the past, with comments on "Radionuclide evidence for low-yield nuclear testing in North Korea in April/May 2010" by Lars-Erik De Geer, *Sci. Global Secur.* **20,** nos. 2/3, 155–171, doi: 10.1080/08929882.2012.711183.

Shearer, P. M., E. Hauksson, and G. Lin (2005). Southern California hypocenter relocation with waveform cross-correlation, Part 2. Results using source specific station terms and cluster analysis, *Bull. Seismol. Soc. Am.* **95,** 904–915, doi: 10.1785/0120040168.

Slinkard, M. E., D. B. Carr, and C. J. Young (2013). Applying waveform correlation to three aftershock sequences, *Bull. Seismol. Soc. Am.* **103,** no. 2A, 675–693, doi: 10.1785/0120120058.

Waldhauser, F., and D. P. Schaff (2008). Large-scale relocation of two decades of Northern California seismicity using cross-correlation, and double-difference methods, *J. Geophys. Res.* **113,** no. B8, doi: 10.1029/2007JB005479.

Wychecki-Vergara, S., H. Gray, and W. Woodware (2001). Statistical development in support of CTBT monitoring, *Defense Threat Reduction Agency Technical Report DTRA-TR-00-22*, 13–39.

Appendix

We used multiplication in the frequency domain to quickly perform the normalized cross correlation between

```
% Megan Slinkard
% Sandia National Labs
% This function does normalized crosscorrelation in the frequency domain.
%
% INPUTS:
% template - vector [Wx1]
% data - vector [D x 1] where W >= D

function [C2] = fastcorrelation(template, data)

figure; plot(template); title('template');
figure; plot(data); title('incoming data stream');

Nd = numel(data);
Nt = numel(template);
N = Nd+Nt; % make sure FFT is properly sized

reversedtemplate = flipud(template); % flip the template in time
xzp = zeros(N,1); %zero pad to make vector length N
xzp(1:Nt,:) = reversedtemplate;
T = fft(xzp);

datazp = zeros(N,1); %zero pad to make vector length N
datazp(1:Nd) = data;
D = fft(datazp);

Z = T.*D; % Multiply in the frequency domain
C2un = real(ifft(Z));  % transform back to the time domain
%unnormalized correlation
%%
%Now find the denominators for normalizing
templateden = sum(template.*template);

rectangle = ones(Nt,1);
rectzp = zeros(N,1);
rectzp(1:numel(rectangle)) = rectangle;

R = fft(rectzp);
D2 = fft(datazp.*datazp);  % square the data signal so that when we do the
convolution we will get the energy in the signal

E = R.*D2; % multiply in frequency domain
dataden = real(ifft(E)); %transform back to time domain

C2 = C2un./((templateden.*dataden).^.5);
C2 = real(C2(Nt:N - 1));  % excerpt segment which corresponds to entire
template being correlated with the raw data.
```

Figure A1. Sample code which implements normalized cross correlation.

the template and a long data stream, in a method well known to electrical engineers. Other resources for frequency domain implementations are Harris and Paik (2006), and section 3.8 of Oppenheim and Schafer (1975). On Dell Precision T7400 workstation, this method was about 200 times faster than the standard time domain implementation.

Let us define the template as $m[t]$ where t goes from $1 : W$, in which W is the length of the window. Let us define the raw incoming data stream $d[t]$ in which t goes from $1 : D$, $D \geq W$ (and often $D \gg W$), and D is the length of the raw data stream. The normalized correlation coefficient for window n can thus be described as

$$\text{corrcoef}[n] = \frac{\sum_{t=1}^{W} m[t]d[t+n]}{\sqrt{\sum_{t=1}^{W} m^2[t] \sum_{t=1}^{W} d^2[t+n]}},$$

in which this can be calculated for all points $n = 0 : D - W + 1$.

A typical time domain implementation would calculate $c[n]$ for each value of n using the above formula. One can think of the numerator, the simple correlation, as convolving $m[-t]$ with $d[t]$:

$$c[n] = m[-t] \otimes d[t+n].$$

In the frequency domain, this is expressed as a simple multiplication:

$$C(f) = M(-f)D(f).$$

By transforming $m[t]$ and $d[t]$ into the frequency domain using the $D + W$ point fast Fourier transform, then multiplying, then transforming back to the time domain, we can calculate $c[n]$ for $n = 1 : D - W + 1$.

Figuring out how to calculate, in the frequency domain, the normalization parameters that make up the denominator of the correlation coefficient can be a bit trickier. The normalization of the template needs to be calculated only once, but the normalization parameter for the long data stream needs to be calculated for each slide of the data window. One way to think of it is as finding the normalization parameters for the long data stream by convolving a rectangle of the length of the template with the squared data stream. In that manner one can figure out what the normalization parameter is for each equivalent of a time domain correlation block.

An example of MATLAB code demonstrating this algorithm is shown in Figure A1.

P.O. Box 5800
Mail Stop 0404
Albuquerque, New Mexico 87108-0404
meslink@sandia.gov
(M.S.)

230B Seismology
61 Route 9W, P.O. Box 1000
Palisades, New York 10964-8000
(D.S.)

Institute of Geophysical Research
4-Chaikinoi Str. Almaty 050020
Kazakhstan
(N.M.)

P.O. Box 5800
Mail Stop 0401
Albuquerque, New Mexico 87108-0401
(S.H., C.Y.)

225 Seismology
61 Route 9W, P.O. Box 1000
Palisades, New York 10964-8000
(P.G.R.)

Manuscript received 18 May 2014;
Published Online 18 November 2014