Using Waveform Correlation to monitor over 3 years of Central Asia

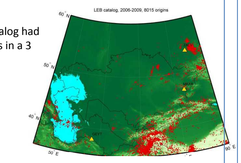
Megan Slinkard*, Stephen Heck*, Dorthe Carr*, Christopher Young* , Paul Richards , Natalya Mikhailova | *Sandia National Laboratories**

SAND2012-10362C

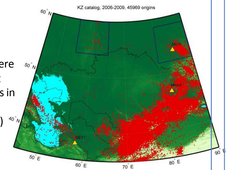
INTRODUCTION

Waveform correlation techniques have garnered increasing attention in the last few years, as their value in detecting and classifying repeated events has been demonstrated again and again. In this research, we show the potential in extending waveform correlation techniques to broad regional monitoring for the benefit of nuclear monitoring. The CTBTO's International Monitoring Service has a sparse network of stations which monitor the globe; we show the ability of waveform correlation to aid catalog completeness.

The LEB catalog had 8015 origins in a 3 year period



The KZ regional catalog had over 45000, AFTER mining events were removed (except for mining events in the two boxed regions in Russia)

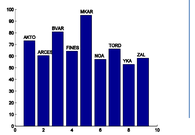


MOTIVATION AND OBJECTIVES

Comparing the CTBTO's LEB catalog to a regional catalog from Kazakhstan which covers central Asia, we note the potential for waveform correlation to enhance the completeness of the LEB catalog.

STATION SELECTION

For this phase of the project, we focused on the benefits of using waveform correlation on just one station. To determine which LMS network station to use for our study, we calculated for each station the percent of events in the LEB catalog (in 2008, in our region) which had Lg arrivals. MKAR was the clear winner, observing 95% of the LEB events (figure 2) . We perform correlation on each element of MKAR's 9 element array.



Percent of LEB events seen at several IMS stations

TEMPLATE SELECTION

Template selection is a critical aspect of a well functioning waveform correlation system. The first question to ask is which phase of an arrival to use as the template. We compared using templates formed from P arrivals, Lg arrivals, and P through Lg in our region of interest. Specifically, we found 99 events with good P and good Lg arrivals; we made templates from these 99 events, set the correlation threshold right at the cusp on consistently getting good matches, processed one year of data, and compared the results. The Lg templates found significantly more matches that P templates; we suspect this is due to their higher signal to noise ratio. P through Lg performed comparable to just using Lg; since just using Lg is computationally more efficient, we decided to use Lg templates for this study.

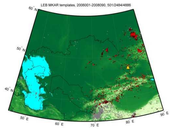
Templates should be clean, clear waveforms, without noise artifacts which will trigger detections on noise segments. We screen templates in a 2 step process; first we require a strong arrival with STA/LTA > 3; second we screen the histograms of correlation values generated using a template - templates which trigger on noise generate many more high correlations that expected and produce unusual histograms. We also use the histogram of correlation values to select a suitable correlation threshold for each template, appropriate to it's time-bandwidth product. Lastly, we cluster the templates and keep only 1 representative template from each cluster.

Query LEB origin and arrival tables for events that meet time, station, and lat/lon requirements

Use STA/LTA to verify observable arrival at station

Verify histogram of correlation values meets histogram criteria (screens noisy templates)

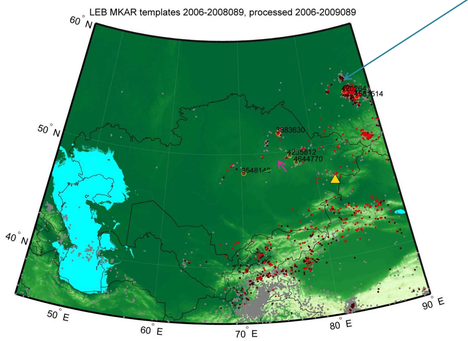
Select correlation threshold (template-specific)



Dates WCD Ran	1/1/2006 – 4/1/2009 (3.25 years) The first 2.25 years overlapped with the period used to make the template library; the last year was processed to study the valued of using archival data for templates.
Stations used	MKAR
Array elements used	9
Templates: acquisition dates	1/1/2006 - 3/30/2008 **2.25 years; 502 templates** Included earthquake swarm in March 2008)
Templates: lat/lon box	lat : 35- 60 lon : 45- 90

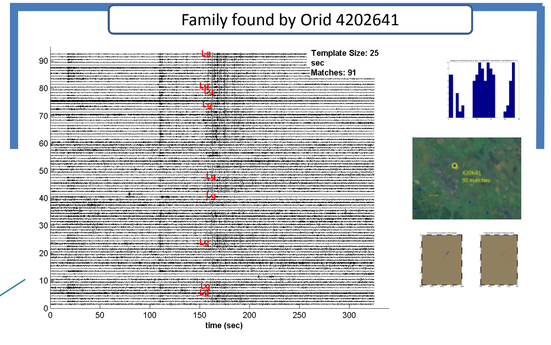
Typical families of similar events are plotted , showing waveform plots and a histogram of the time of day at which the events occurred (to help distinguish minining families from earthquake families). In the waveform plots the top waveform is the template; below it are detected events, sorted by correlation value. The first detection is always the template finding itself; this serves as a nice sanity check, and is not counted in our detection statistics.

Geographic Distribution of Waveform families

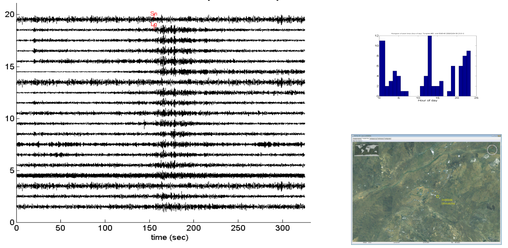


RESULTS

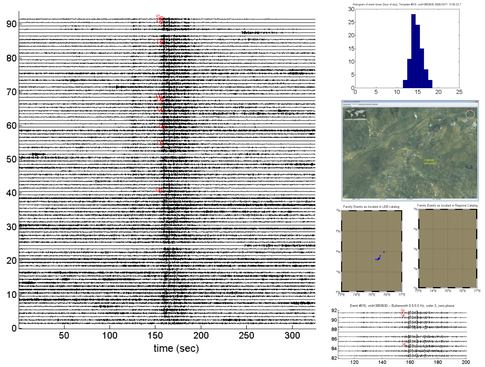
Family found by Orid 4202641



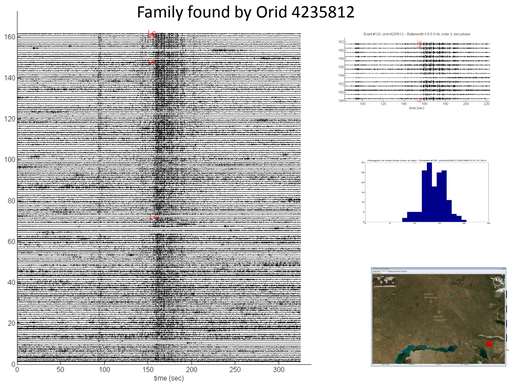
Family found by 3548148



Family found by Orid 3883630

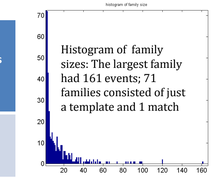


Family found by Orid 4235812



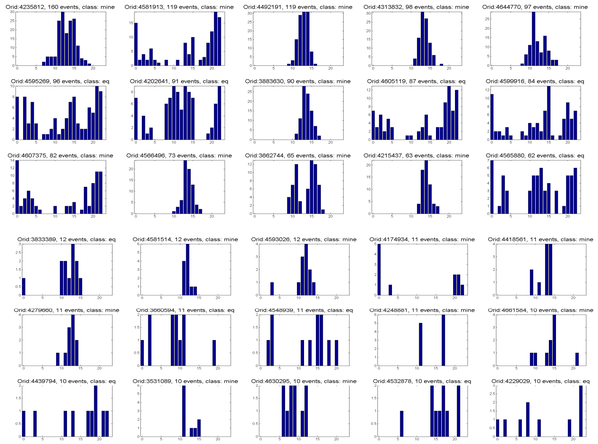
RESULTS SUMMARY

Number of Detections (made by 180 of the 501 templates)	Detections in LEB Catalog	Detections in Regional Catalog *(had mining events removed, except for in the boxed regions!)*	Detections validated in EITHER catalog	# Detections thought to be Mining / earthquakes
5364	429 (2131 events in the LEB catalog)	450	659	>2952 / <1399



Histogram of family sizes: The largest family had 161 events; 71 families consisted of just a template and 1 match

Histograms of the hour of day during which events occurred help us classify families as mining or earthquake families. Our large families were generally classified as mining explosions.



SUMMARY

We performed waveform correlation on 3 years of data for station MKAR in central asia using 501 templates. Our computational abilities allowed us to perform this analysis in 2.5 days. These 501 templates detected 5364 events, of which all but 429 were not in the LEB catalog; this illustrates the value of waveform correlation in aiding catalog completeness. A surprising number of events were not in the regional KZ catalog; this is due to the fact many of our detections were mining events, which had been carefully screened out of the KZ catalog.

FUTURE WORK

Bigger template library – we can only detect what is in our library, so the bigger the better
- Could use a longer period of the LEB
- Could try to develop a library based on the KZ catalog

Expand to other stations
- Will find additional events, because this will decrease our event to station distance for some events
- Also can be used to corroborate events – currently, the threshold for event detection is fairly subjective and unless it is set very high, you will probably include some events that won't meet analyst scrutiny. We could come up with a metric that says that once you are below a certain correlation value, you must correlate at more than one station.

Bigger area – our ultimate goal is to try to do global monitoring

* {meslink, dbcarr, sheck, cjyoung}@sandia.gov // Ground-Based Nuclear Explosion Monitoring Research and Development