

Waveform Correlation Effectiveness During High Analyst Workload Sequences

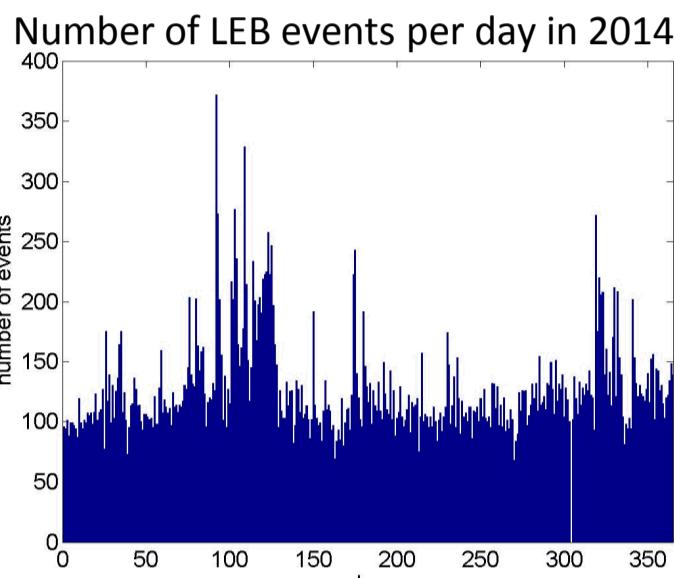
Megan Slinkard*, Stephen Heck, Nedra Bonal, David Daily, Christopher Young | *Sandia National Laboratories*

INTRODUCTION

Waveform correlation techniques have garnered significant attention the last few years, as their value in detecting similar events has been demonstrated again and again. In this research, we show the potential for waveform correlation to aid IDC analysts during their times of highest workload. We estimated analyst workload for all events in the late event bulletin (LEB), and identified source regions and time periods that required significant analyst effort. We then performed waveform correlation at nearby stations and examined how waveform correlation could have aided detection.

MOTIVATION

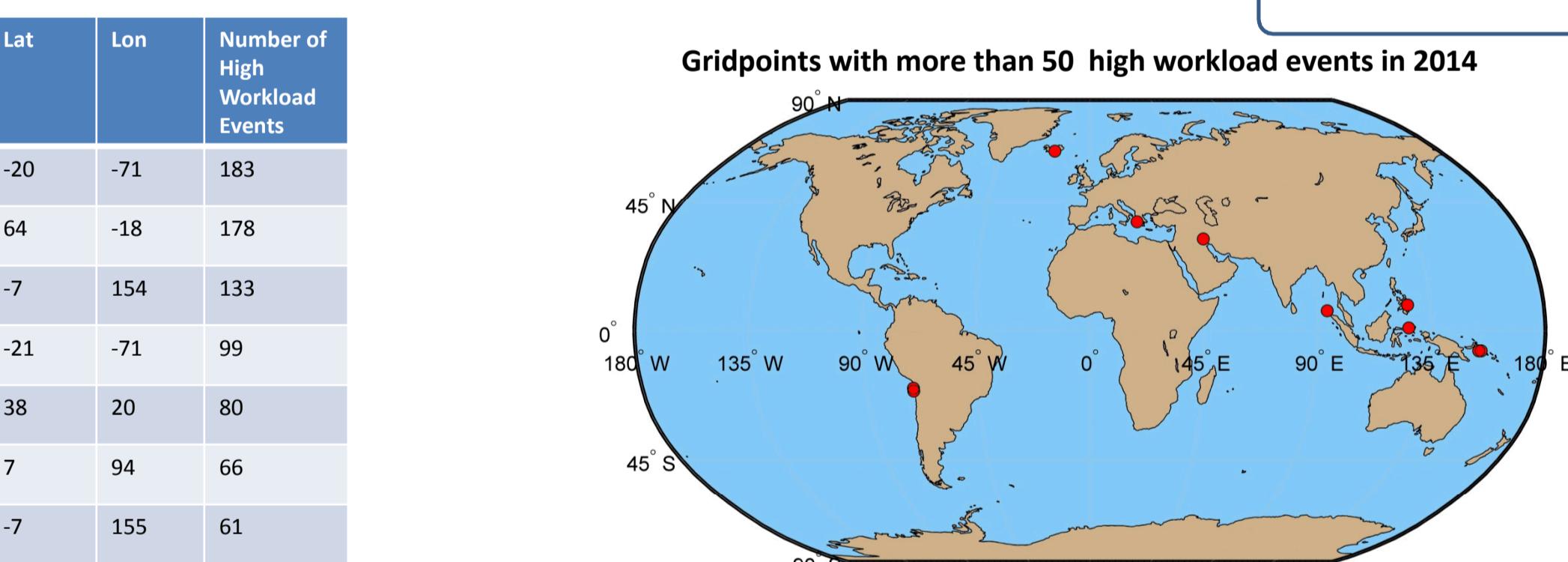
Time distribution of earthquake is highly non-uniform; for a monitoring agency charged with looking at all events in a timely manner, the variance in number of events per day means variance in workload. When major aftershock sequences occur, the workload for analysts can increase dramatically.



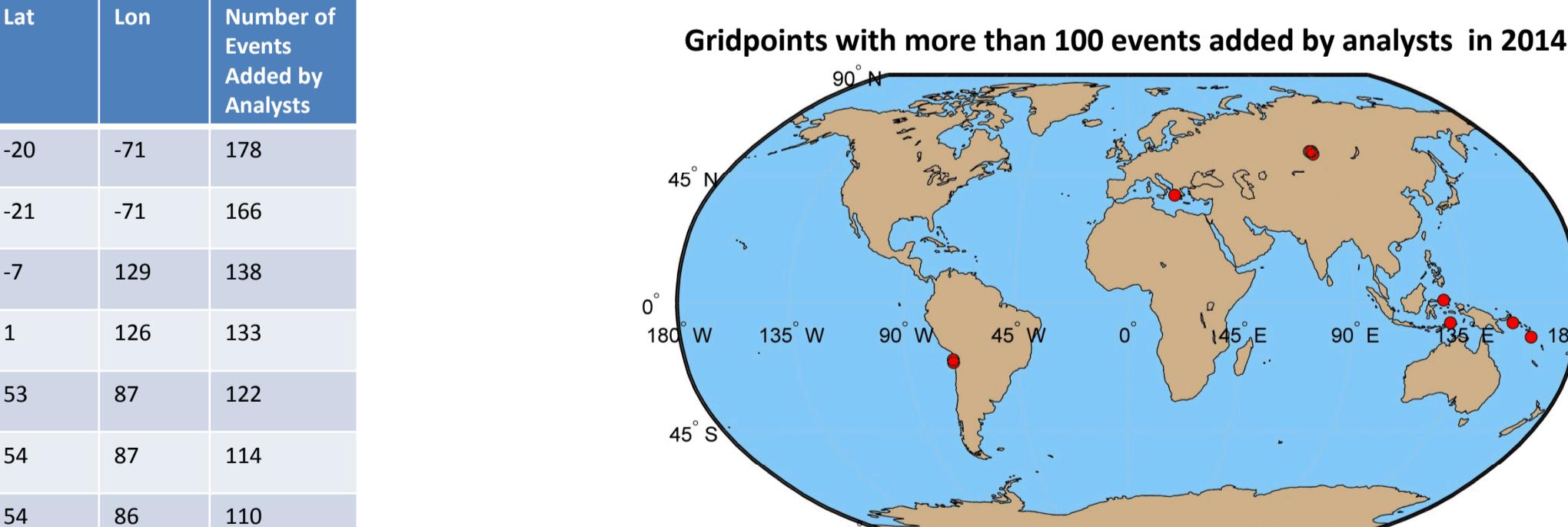
A typical day sees approximately 100 events, but in 2014 on occasion the number of events spiked to almost 4x that.

Not all events take equal analyst time to process. Some events formed by the automatic system and listed in the final automatic event list, the Standard Event List 3 (SEL3), pass to the LEB with minimal changes. Others require significant reworking by the analysts. Moreover, some events are not formed by the automated system but are seen and built by analysts. We looked at all the events in the LEB bulletin in 2014 and identified regions of the world where significant analyst effort had gone into creating the bulletin.

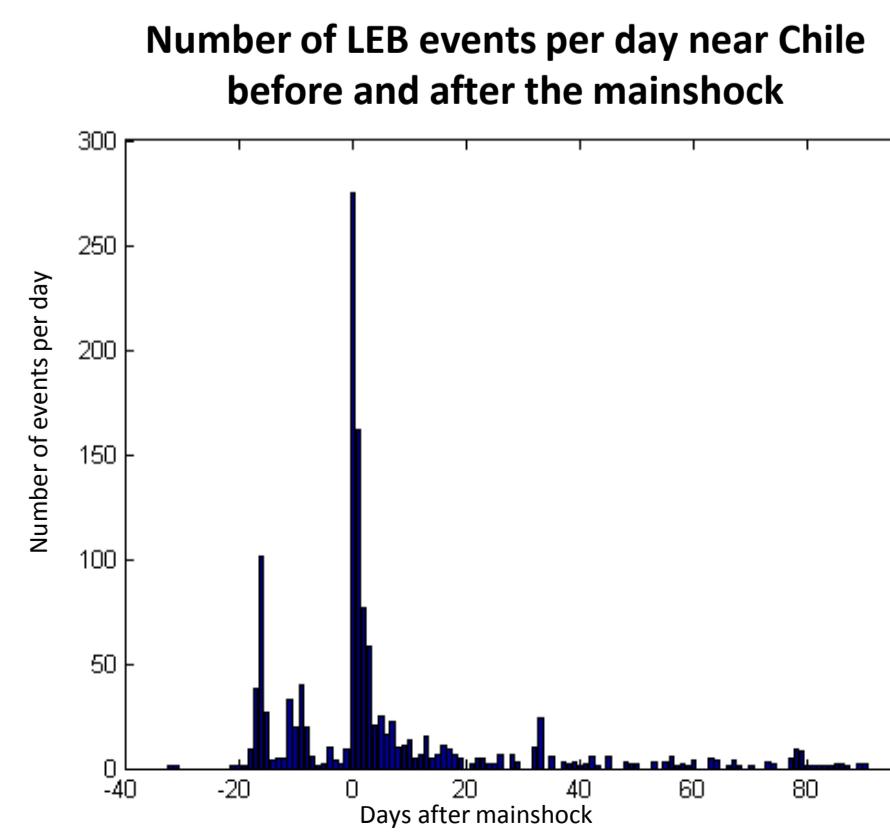
First we ranked events that were listed in both SEL3 and LEB according to our algorithm for evaluating workload (see sidebar). The 20% of events with the highest workload were deemed "High Workload Events". Below we plot the number and ratio of high workload events for $1^\circ \times 1^\circ$ gridpoints around the globe and show a table listing the number of high workload events in the most active gridpoints. The region of the globe with the largest number of High Workload events was off the coast of Chile.



Second, we looked for events that were in the LEB but not in SEL3, as these were noticed and built by an analyst (a much more time consuming process than confirming events built by the automated system). We called these events "analyst-added". We plot the number of analyst-added events for $1^\circ \times 1^\circ$ gridpoints around the globe and list the information for the most active gridpoints in the table. The region of the world with the most number of analyst-added events was again off the coast of Chile.



REGION OF STUDY



As the region off the coast of Chile had the largest number of high workload events and analyst-added events, we use this region in our study. There was a large magnitude 8.2 earthquake off the coast of Chile on April 1, 2014. A magnitude 7.6 aftershock followed on April 3. Most of the high workload and analyst-added events occurred during that aftershock sequence.

A small foreshock sequence occurred in the 30 days prior to the mainshock (March 1-31). We used these events as templates and studied our performance detecting events during the main aftershock sequence (April 1 – July 1).

EXPERIMENTAL SETUP

The aftershock sequence that followed the April 1, 2014 earthquake off the coast of Chile resulted in a significant workload increase for the analysts. We evaluated the performance of waveform correlation during the 3 months after the mainshock. For template events we used events from the previous month (March 1-31). The closest primary station was PS06 (LPAZ) (4.2° from mainshock); two other nearby primary stations were PS30 (CPUP, 14.1°) and PS01 (PLCA) (21.0°).

Dates SeisCorr ran	April 1–July 1, 2014 (mainshock occurred April 1, 2014)
Stations	LPAZ, CPUP, PLCA (IMS Primary Network)
Components	BHZ, BHN, BHE
Template Dates	3/1/2013 – 3/31/2014
Template Lat/Lon Box	[-22° : -18° , -73° : -69°]
Template filtering	.5 – 10 Hz, 3 rd order Butterworth
Template window size	30 seconds, starts 5 seconds before arrival

SEISCORR software

SeisCorr performs 3 main functions: 1) Creation of a Library of template events, 2) Waveform Correlation processing, 3) Detection Validation. Its graphical user interface also allows viewing of templates, detections, validation results, and associated maps and metadata.

A template library consists of waveform segments of known events that will be used to search for similar events. For optimal results, templates should be good quality, high signal-to-noise ratio waveforms. SeisCorr allows the automatic screening of template candidate events; however, screening can result in the accidental omission of useful templates. In this case, we used all events in the LEB that had an arrival at the station as templates. Each template was then assigned a correlation threshold based on a desired false alarm rate of 1 per month. Our correlation thresholds were determined using our reverse-template method.¹

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. For large datasets this is computationally expensive, hence SeisCorr runs distributed on multiple computers to allow for quick processing.

Detection Validation is done as a post-processing step, called Multi-Station Validation, where detections at each station are compared.

RESULTS

The Chilean aftershock sequence (as defined by our latitude/longitude and time box) had 952 events listed in the LEB bulletin. Of these, 192 were considered "high work", and 343 were added by hand.

For each of the 3 primary stations in the region, we used a template library filled with foreshocks from the previous month and did waveform correlation to detect similar events. We then evaluated how many of the LEB events and high workload events with arrivals at that station were detected.

At station LPAZ, Waveform Correlation was able to detect a significant portion of the LEB origins (28%) in the aftershock sequence. Moreover, it detected 55 events that had to be added by hand, and 82 (44%) of the events labelled high workload.

Station	LPAZ
Template Library	401 events
Number of LEB events in aftershock sequence with arrivals at LPAZ	882
Detections: total number	1456
Detections: in LEB	248 (28% of LEB events with arrivals)
Detections: added by hand	55 (19% of arrivals added by hand)
Detections: labelled high workload	82 (44% of high workload arrivals)

The other two stations were much farther from the mainshock and did not perform nearly as well – they had far fewer detections in general, and only 18 and 5 (CPUP and PLCA, respectively) were linked to LEB events.

Station	CPUP (14.1° from mainshock)
Template Library	154 events
Number of LEB events in aftershock sequence with arrivals at CPUP	321
Detections: total detections	101
Detections: in LEB	18 (6% of LEB events with arrivals)
Detections: added by hand	0 (0% of arrivals added by hand)
Detections: labelled high workload	17 (10% of high workload arrivals)

Station	PLCA (21.0° from mainshock)
Template Library	181 events
Number of LEB events in aftershock sequence with arrivals at PLCA	293
Detections: total number	55
Detections: in LEB	5 (<1% of LEB events with arrivals)
Detections: added by hand	0 (0% of arrivals added by hand)
Detections: labelled high workload	5 (17% of high workload arrivals)

Multi-station Validation Results

15 event were seen at 2 or more stations, 3 events were seen at all 3 stations.

Details of Results Analysis:
"Total detections" lists the number of detections (whether or not we linked the detection to an LEB event). As the plots indicate, many of these waveforms appear to be from events very similar to the template event.
Determining whether a detection was an LEB event was done by determining the travel time associated with the template, subtracting that travel time off the detection time to form a calculated origin time, and seeing if there was an LEB event within a 4 second window. Sometimes an LEB event was detected both from a P template and an Lg template; only unique LEB detections are listed.

SUMMARY AND DISCUSSION

The Chilean aftershock sequence (as defined by our lat/lon and time box) had 952 events listed in the LEB bulletin. Of these, 192 were considered "high work" based on the number of re-pickings and recalculation required significant analyst time. Using template events from the month *before* the mainshock we were able to detect a significant portion of events in the aftershock sequence. Station LPAZ was the closest station and had the best performance; we detected 248 (26%) events listed in the LEB bulletin. SeisCorr did a particularly good job of detecting events we identified as "high workload" – we detected 82 (42%) of those. Moreover, SeisCorr detected 55 events that had been added by hand. We also found numerous good detections not listed in the LEB.

The other two closest primary stations, PLCA and CPUP, did not perform nearly as well, however, and only 15 events were detected at two or three stations. These two stations had fewer arrivals during March from which to build templates, and these templates were often of very low SNR. Although waveform correlation techniques can do a fantastic job of detecting a signal in noise, that requires a good SNR template with which to start.

The results at LPAZ, however, suggest that incorporating Waveform Correlation processing into the processing flow could be of value during times of high analyst workload. A related poster ("Improved Bulletin Generation Using An Iterative Processing Framework") discusses a pipeline scheme which incorporates SeisCorr detections and feeds them to the associator, along with their associated metadata (calculated origin time, estimated lat/lon, estimated magnitude, etc.). Such a scheme could lead the automated system to build the events that were previously build by hand, and lead to better initial characterization of the high workload events. Knowledge of the additional events detected by SeisCorr could lead to a more complete bulletin; moreover, even in the case where the additional detections are too low magnitude to be included in the final bulletin, knowledge of them would still help the associator properly build events.

FUTURE WORK

- Use Template libraries with results from first day or so of the aftershock sequence, and see how much that improves performance.
- Add events to the Template Library as SEL3 builds them and see how that improves performance.
- Explore additional regions which ranked high on our High Workload and Analyst-Added lists.