SANDIA REPORT
SAND2020-10548
Printed September 2020

Sandia National Laboratories

Applying Compression-Based Metrics to Seismic Data in Support of Global Nuclear Explosion Monitoring

Laura E. Matzen, Christina L. Ting, Richard V. Field, James D. Morrow, Ronald Brogan, Christopher J. Young, Angela Zhou, Michael C. Trumbo & Jamie L. Coram

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

The analysis of seismic data for evidence of possible nuclear explosion testing is a critical global security mission that relies heavily on human expertise to identify and mark seismic signals embedded in background noise. To assist analysts in making these determinations, we adapted two compression distance metrics for use with seismic data. First, we demonstrated that the Normalized Compression Distance (NCD) metric can be adapted for use with waveform data and can identify the arrival times of seismic signals. Then we tested an approximation for the NCD called Sliding Information Distance (SLID), which can be computed much faster than NCD. We assessed the accuracy of the SLID output by comparing it to both the Akaike Information Criterion (AIC) and the judgments of expert seismic analysts. Our results indicate that SLID effectively identifies arrival times and provides analysts with useful information that can aid their analysis process.

CONTENTS

This page left blank

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
AIC	Akaike information criterion
NCD	Normalized compression distance
NID	Normalized information distance
SLID	Sliding information distance

1. INTRODUCTION

Seismic analysis is an important component in global nuclear treaty monitoring. Analysts comb through seismic data collected by the International Monitoring System (IMS) global seismic sensor network to help verify compliance with the Comprehensive Nuclear Test-Ban Treaty (CTBT). They make "picks" to indicate the times at which seismic signals (i.e., transient vibrations) from events of monitoring interest reached different sensors around the world. It is important for the timing of the picks at each sensor to be as accurate as possible in order to detect and characterize the corresponding events. The timing information taken from multiple seismic stations is used to calculate the latitude, longitude, and depth of the event that produced the seismic signals. In turn, the location information helps determine whether the event was most likely to be an earthquake or a man-made event such as a nuclear test.

Seismic analysis in support of nuclear treaty monitoring is a very important task that can also be very challenging. The analysis process for CTBT monitoring begins with data collected from the 170 seismic stations that are part of the IMS. These data are sent to the International Data Centre (IDC) at the headquarters of the Preparatory Commission for the Comprehensive Nuclear Test-Ban Treaty Organization (CTBTO), where they go through an automated station processing stage that produces a list of the seismic signals identified at each station. These lists form the input for three rounds of automated network processing, which associates the signals with one another across the whole network to identify seismic events. The result is the Standard Event Lists (SELs). The third iteration (SEL3) is reviewed by human analysts who clean up the automatic detections to produce the Reviewed Event Bulletin (REB).[1] On average the analysts reject about 40% of the automatically built events and manually identify new events that were missed by the automated processing. These new events make up an average of 20% of the events in the REB.

The seismic analysts' work is both tedious and stressful. Each waveform is "touched" by multiple analysts who have different subjective preferences about how to place picks. The data are often noisy, with overlapping events and background noise that make it difficult to pinpoint event arrival times. Yet the pressure to produce a reviewed event catalog as quickly as possible is unrelenting and the stakes are high; ideally, all of the potentially numerous seismic signals on each sensor in the datasets must be accounted for, to rule out the possibility that any one of them was caused by a nuclear test.

In the project described in this report, our goal was to develop a new approach to analyzing seismic waveforms for signals, with the intention of providing analysts with information that will make a difficult job a little bit easier. We developed new methods for applying compression-based metrics to seismic waveforms in order to detect boundaries, indicating the arrival of a seismic signal at a particular station. Our approach was inspired by previous research on using compression-based metrics as a measure of analytic completeness in text analysis. By applying these same metrics to waveforms using a moving window approach, we were able to identify event arrival times with accuracy that was equal to or better than the current state-of-the-art algorithms that are used for automatic event detection. In addition, we demonstrated that providing visual representations of the compression scores to analysts provides useful cues that can help them with making their own picks.

[1] See https://www.ctbto.org/press-centre/news-stories/2001/data-processing-at-the-international-data-centre/ for more details about this process.

In the sections that follow, we will describe the two compression-based metrics that were used in our research: Normalized Compression Distance (NCD) and Sliding Information Distance (SLID). We will outline the moving window technique that was developed to apply these metrics to waveforms and describe the results obtained from applying these metrics to two datasets of seismic events. The output of the compression-based metrics was compared to picks made by an expert seismic analyst as well as the output from the Akaike Information Criterion (AIC), the algorithm that is currently used for refining the timing of automatic picks before analyst review. We also demonstrate that an uncertainty quantification (UQ) approach applied to SLID may be used to rank the confidence in our event predictions, thereby providing a ranking for which to present the SLID output to analysts. Finally, we will discuss the design and results of a user study in which different visual representations of the SLID output was presented to seismic analysts to aid in picking signals from waveforms.

1.1. Background on Compression Metrics

Compression-based metrics such as the normalized compression distance (NCD) involve applying a compression algorithm for assessing the similarity of any two items (cf. Li et al., 2004; Vitányi et al, 2009). Compression algorithms reduce the number of bits to encode an item by assigning fewer bits to more common patterns[2]. The same principle can be applied to compare two items and this property forms the foundation of compression-based metrics. Specifically, two items belonging to the same class will have more similar patterns and better compression than two items belonging to different classes.

There has been extensive prior research at Sandia in applying compression-based metrics such as NCD in a variety of applications. For example, one prior study used NCD to distinguish between truthful and deceptive documents (Ting, Fisher & Bauer, 2017). The basic assumption is that truthful documents should share similar patterns and deceptive documents should share similar patterns distinct from truthful documents. The NCD is used to assess similarity.

Building on these principles, other research at Sandia has successfully applied NCD as a metric of the analytic process in a variety of contexts. McNamara and colleagues (2016) demonstrated that NCD could be used as a measure of analytic progress for large sets of text documents. In that study, participants were given a task that mimicked an intelligence analysis task. They were given a large set of documents (200 total) with meaningless file names and were asked to categorize them and to describe any noteworthy discoveries they made in the process of analyzing the documents. The noteworthy information in the documents was about Alexander Litvinenko's death, but this information was buried beneath a pile of unrelated or tangentially related information. They hypothesized that the value of NCD would decrease as participants organized the documents into topical groups. The documents about similar topics should contain similar words, so increasing the organization of the information and grouping documents according to their topics should reduce the entropy within each group and therefore decrease the value of NCD. The study confirmed that NCD scores decreased as participants organized the documents into coherent categories. An example is shown in Figure 1, which shows the number of categories that this participant created over time and the mean NCD score for the dataset. As the number of categories increased and the

[2] For example, if one sees a *q* in a sequence of English text, the likelihood that the next character is a *u* is so high that it is almost not necessary to include the *u*. Therefore, it is advantageous to use fewer bits to represent *qu* than to represent something like *q-*.

categories were refined to contain sets of documents that were more similar to one another, the mean NCD score for the dataset decreased.

This work, which demonstrated that NCD could track with analytic progress inspired the current project. Much of the work of seismic analysis stems from refining automatically built events. The analysts reject extraneous picks and signals that were most likely produced by a different event. They adjust pick times and add picks when necessary, with the goal of building an event that combines data from multiple seismic stations and produces an estimate of the location and magnitude of the seismic event that is as accurate as possible. Conceptually, sorting seismic signals from different stations into coherent events is similar to organizing a pile of text documents into coherent categories. We hypothesized that NCD could provide a measure of whether changes in the seismic analysis, such as moving picks, were helping or hindering the goal of building a coherent event.

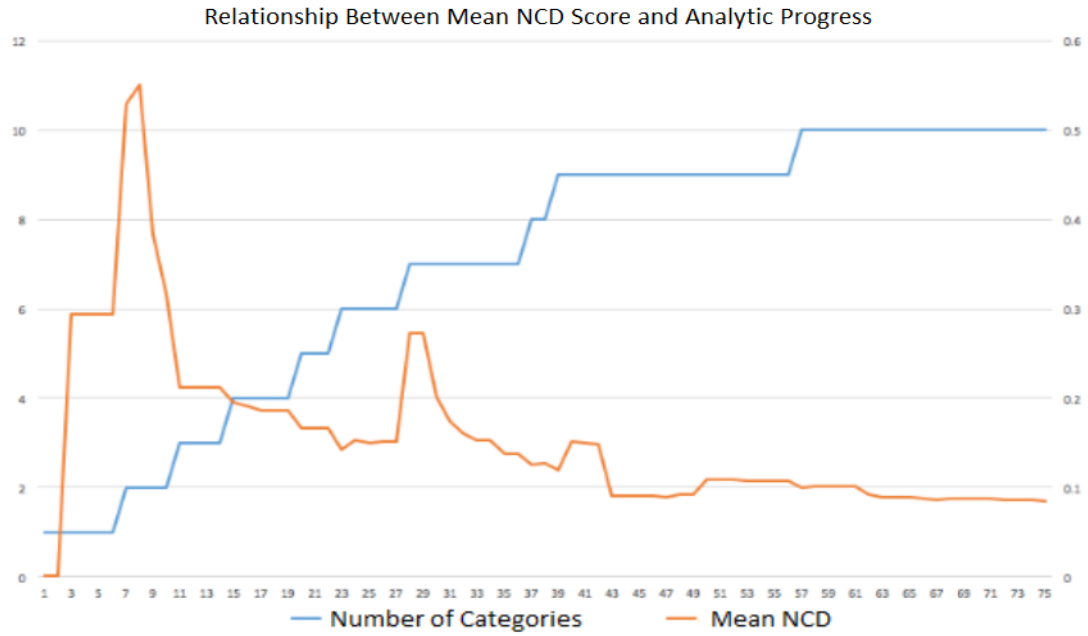

Figure 1. The relationship between analytic progress and the mean NCD score for a set of text documents.

Our first step was to determine whether compression metrics such as NCD could be adapted to detect transitions from one state to another in seismic waveforms (i.e., the arrival of seismic signals). As described below, a common approach for detecting changes in sequential data is to move a sliding window along the data and compare adjacent windows. In the case of NCD, the NCD scores should peak when the two sections being compared are maximally different from one another. In other words, the maximum NCD value should correspond with the event arrival time. If that maximum value corresponds closely to the picks made by an algorithm or analyst, it indicates that the pick is well-supported by the information contained in the waveform and that there is no need for further adjustment of that pick. Similarly, when applied across multiple waveforms from multiple seismic stations, the NCD peaks for seismic signals generated by the same event should align with the expected arrival times based on the distance between the event and the station. In cases where the NCD peak does not align with the expected arrival time, it is likely that the corresponding signal originated elsewhere and does not fit with the event being analyzed.

In the sections below, we will discuss how NCD was calculated for seismic waveforms and how the results compared to the current state-of-the-art algorithm, AIC, as well as expert judgements of signal arrival times. To put the bottom line up front, we found that there is a clear NCD peak for the majority of the waveforms analyzed and that it corresponds very closely to the picks made by AIC and by an expert analyst. In addition, plotting the NCD values over time provided useful information about when expert judgments were needed or when a different filter on the waveform would provide better information to analysts. The primary drawback of NCD is that it is fairly slow to calculate, which could be problematic for real-time analysis. To address this drawback, we tested another variant of the compression-based metrics called the sliding information distance (SLID). We found that SLID and NCD produced similar results, but the SLID calculations were much faster. The speed of the SLID calculations enabled us to calculate SLID multiple times for each waveform with varying parameters. This approach provides a quantification of the uncertainty in the data by assessing a large parameter space.

2. APPLYING THE NORMALIZED COMPRESSION DISTANCE TO SEISMIC WAVEFORMS

Let $z = \{z_0, z_1, \ldots\}$ be a time series representing a seismic waveform. Identifying seismic events in z requires significant latency and domain knowledge of the underlying data. To reduce the time and expertise cost, *change-point (i.e., event) detection* methods can be applied to automatically determine events in z.

General change-point detection methods often involve sliding adjacent windows over sequential data (here the seismic waveform), collecting statistics of the underlying data within each window, and computing a distance function that operates on the statistics to determine large distances between adjacent windows (Aminikhanghahi & Cook, 2017). Large distances that manifest as *peaks* are then flagged as *events* in the seismic waveform.

2.1. Calculating the Normalized Compression Distance

One approach to change-point detection is to use a distance function based on the *normalized information distance* (NID). The NID between any two items, x and y, represents ``the minimal quantity of information sufficient to translate between x and y, generating either item effectively from the other'' (Vitányi et al., 2009). In our application, x and y refer to adjacent windows in the seismic waveform, as represented by the shaded boxes in Figure 2.

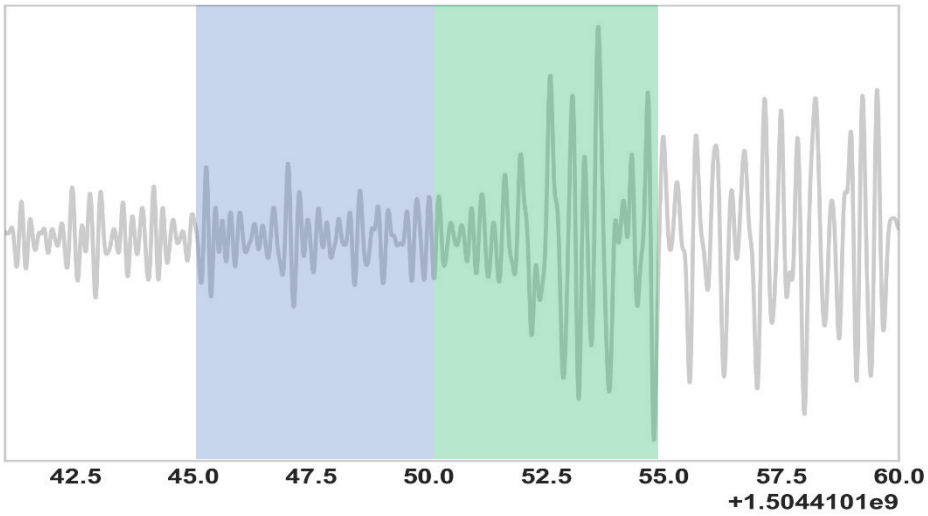

Figure 2. A seismic waveform where x is represented by the blue shaded region and y is represented by the green shaded region.

The NID is defined by

$$NID(x,y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}},$$

and takes values in $[0,1]$. In this expression, $K(x)$ denotes the Kolmogorov complexity of x and is defined as the length of the shortest possible description of the sequence in some fixed universal description language (Li & Vitányi, 1997). Similarly, $K(x|y) \leq K(x)$ represents the length of the shortest possible description of x when y is provided as auxiliary information. For general x the

complexity $K(x)$ is not a computable function, so that approximations for the NID are needed in practice. Accordingly, there has been extensive work on approximating the NID, most of which are based on the normalized compression distance (NCD) (Li et al., 2004).

The NCD is defined by

$$NCD(x,y) = \frac{|\mathrm{C(xy)}| - \min\{|C(x)|, |C(y)|\}}{\max\{|C(x)|, |C(y)|\}},$$

where $|C(x)|$ denotes the compressed size of x after applying a compression algorithm C, and xy denotes the concatenation of items x and y. Approaches using the NCD have proven successful in a variety of machine learning applications. Examples include authorship attribution (Stamatatos, 2009), image registration (Bardera et al., 2006), evolutionary history inference (Li et al., 2004), and cybersecurity (Wehner, 2007).

2.2. Applying NCD to Seismic Waveforms

While NCD has been shown to be useful in a variety of domains, it had not previously been applied to seismic waveform data. To test our approach, we first created synthetic seismic waveforms to provide "ground truth" for evaluating the NCD results. The synthetic signal has the form:

$$s(t) = \mathrm{A}\exp\left(-\frac{5(t-t_0)}{T}\right)\cos(2(t-t_0))\sin(\omega(\mathrm{t}-t_0)+\phi), \qquad t \geq t_0$$

It begins at $t = t_0$ and so has a maximum amplitude of A, oscillates as a cosine with frequency $\omega > 0$, and exponentially decays at a rate proportional to parameter $T > 0$. Two such signals were added together, with a smaller amplitude signal shifted to the middle of the time array ($t_0 = 15$), as shown in the top panel of Figure 3. This represents a common scenario in seismic analysis, where signals of interest may appear close in time to larger, unrelated signals. Then, to simulate the background noise present in seismic data, iid Gaussian noise with zero mean and standard deviation equal to 0.1*A was added to this signal, as shown in the bottom panel of Figure 3.

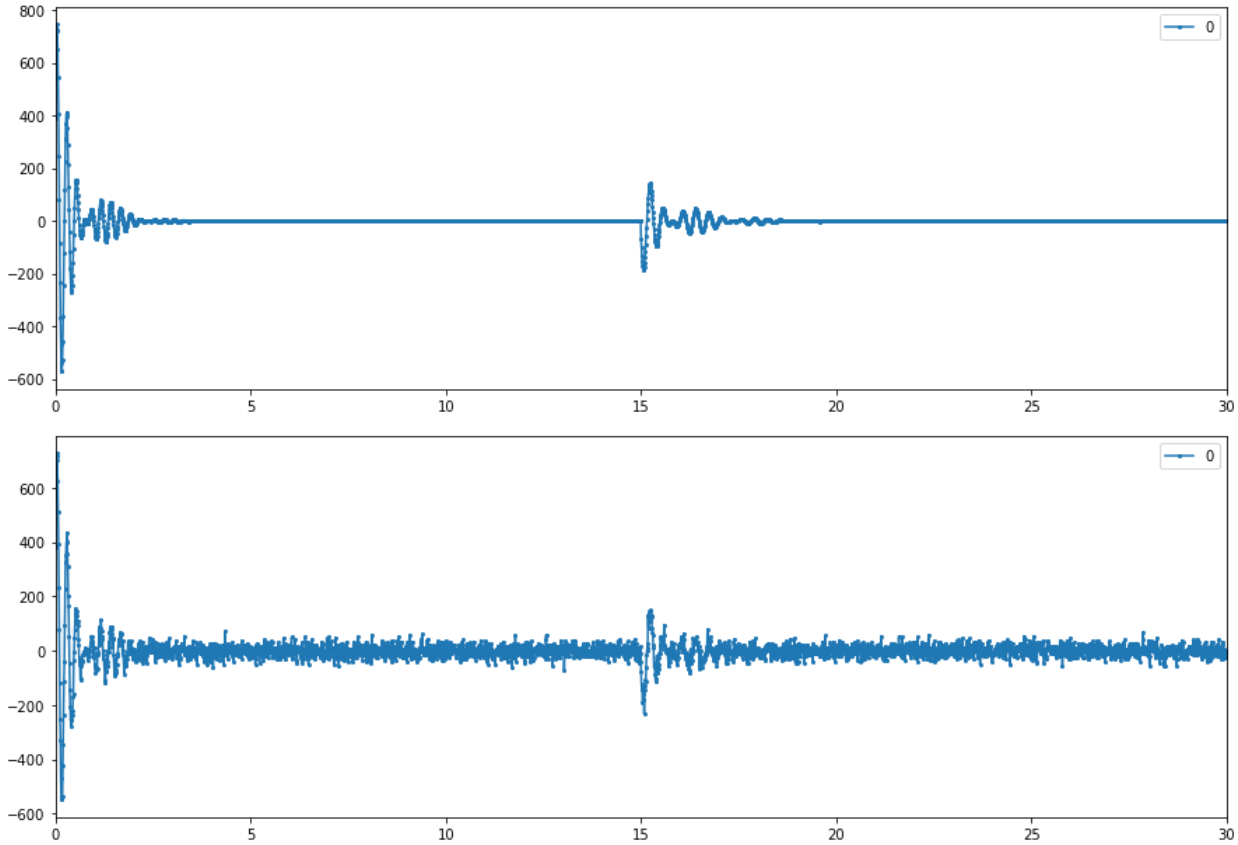

Figure 3. The superposition of two synthetic seismic waveforms (top) and the synthetic waveform with iid Gaussian noise added (bottom).

2.2.1. Computing an NCD Value

Given a time series of real-valued data, we would like to compute a series of NCD values for this series. The first thing we need to do is tokenize the data – discretize it using 256 tokens (8 bits)[3]. The next step is to move a sliding window along the data. The NCD is computed on a window by subdividing the window into a left-half and right-half. Each window position yields a single NCD value that represents the comparison between the two halves. To the extent these halves are very statistically similar in terms of their values, the NCD score will be low – indicating a "small" distance between them. Note that even if the two halves are identical, the NCD score will not be zero, just "small". Likewise, if the two halves are very statistically dissimilar, then the NCD score will typically be 1 or higher, but there is no absolute maximum value.

There were several key issues that needed to be considered in developing this approach. The first is the impact of discretizing all of the data at once versus discretizing at each window location. Discretizing all of the data before applying the moving window is simpler and saves computation time, but it could cause us to miss smaller (in amplitude) seismic signals when there are larger, unrelated events in the same data segment. Discretizing within each window could improve the dynamic range of the calculation, particularly when the range or values varies widely along the time series. A second important issue is the width of the windows. If the window is too small, there is not enough data from which to estimate the statistics. If the window is too large, computation time

[3] Other numbers of tokens are possible, but we find that 256 works well in practice.

increases and different features in the signal may become conflated. Finally, we must consider how much to slide the window between calculations. Moving the window too much could make it difficult to locate the peak NCD value for the waveform, but moving it too little slows down the analysis considerably because NCD is calculated many more times[4]. These three key issues are discussed in more detail in the following three sections.

2.2.2. Discretizing Data for NCD Calculation

We require the real-valued sequence of values that makes up a seismic waveform to be discretized to tokens (0-255) in order to compute NCD score for a window. There are two basic ways to approach this. The standard approach is to discretize all of the data in the series at one time, and then slide the window along, computing an NCD value for the midpoint of each window position. One potential disadvantage of this approach arises if the signal has most of its variation in values in a small part of the signal. In this case, much of the signal might be represented by only a few token values. A second approach is to discretize at each window location. This may maximize the signal-to-noise ratio (SNR) because a maximal number of token values is used to represent the signal within the window. The disadvantage is additional computation at each window location.

To assess the utility of each approach, we calculated NCD scores for a series of synthetic seismic waveforms with different levels of background noise. Like most real seismic waveforms, the synthetic waveforms were sampled at 100 Hz. The waveforms were either discretized all at once, or window-by-window. NCD scores were calculated using a window width of 8 seconds (800 samples) and a window step size of 0.8s (80 samples). The resulting NCD scores were scaled from 0-1 to make comparison easier. These comparisons indicated that discretizing the waveform window-by-window produced greater dynamic range and better SNR. The per-window NCD values were generally either higher or lower than the NCD values computed using the entire signal discretization (see Figure 4 for examples). However, the locations of the maximum NCD peaks were nearly identical for both methods. As the waveforms grew noisier, the per-window discretization did not provide a substantial benefit above and beyond the whole-waveform discretization. Given the extra computation required to discretize each window separately, we determined that there was not a significant advantage to using the per-window approach. n the remainder of this report, we discretize the entire waveform in our analyses. Figure 4 shows examples of the per-window and whole-waveform discretization approaches for synthetic signals with varying noise levels as well as a real seismic signal.

[4] We note that SLID (described in the following Section) is computationally efficient and eliminates this last issue so that we simply move the window every time step.

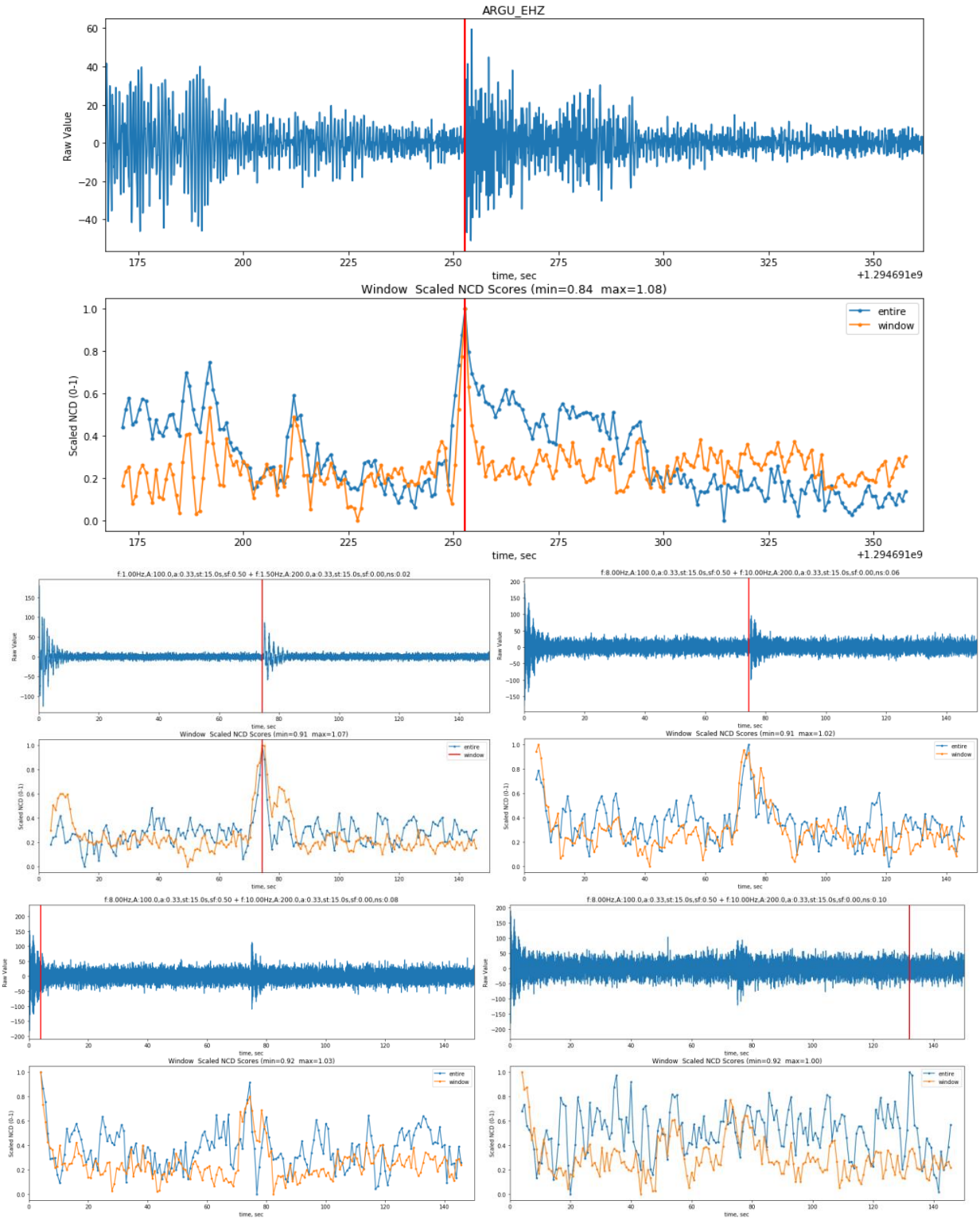

Figure 4. The top section shows a real seismic signal above the scaled NCD values that were calculated using either the entire waveform (blue line) or per-window discretization (orange line). For both discretization methods, the maximum NCD value corresponds to the event arrival time (indicated with red lines). The bottom section shows these same calculations for four synthetic seismic signals with increasing levels of background noise. The NCD peaks align for both methods, and both fail to detect the signal of interest when the noise level gets too high.

2.2.3. Effects of Different Window Sizes

Window size is a very important parameter for the NCD calculations. The window must be wide enough to capture representative data but narrow enough to effectively identify transition points in the waveforms. Importantly, the window width should account for the frequency content in the signal. If the window is too short with respect to the frequency content, the resulting NCD values will oscillate in a way that could hinder detection of an event transition. The sampling rate of the data is also important. We need to have enough samples in each half of the window to generate reasonable statistics. Since most seismic data is sampled at 100 Hz, a 2-second window is likely to be too short, as it would only have 100 samples in each half of the window. A 12-second window would have 600 samples in each half, which would allow us to generate robust statistics, but risks conflating separate seismic signals by including them in the same window.

To assess the impact of window size on the NCD results, we calculated NCD values for our synthetic waveform and for real seismic waveforms using different window widths. The window was shifted by 100 samples for each calculation. An example of the results of this analysis is shown in Figure 5 for window sizes of 2, 4, 8, and 12 seconds. When the window width is larger, the NCD scores reach a higher absolute peak. As expected, there is less noise for the larger windows because there is more overlap between one window and the next. The vertical green lines in the figure represent the earliest, most likely, and latest possible arrival times for the seismic signal, as determined by an expert analyst. The 12-second, 8-second, and 4-second windows all produced a maximum NCD value that fell within the range indicated by the expert analyst. The 2-second window produced several peaks with similar heights, with the maximum value occurring about 10 seconds later than the actual signal onset. This analysis indicated that a window width should generally be 4 seconds or more to produce a stable result for seismic data that is sampled at 100 Hz.

Another consideration in choosing the window width is the computation time. We analyzed the median time it took to calculate NCD values when different window sizes were applied to the same seismic waveform. We found that there was a linear relationship between window width and computation time, as shown in Figure 6. From a computational point of view, this favors using a window that is as narrow as possible. Interestingly, we also observed that the windows with the longest computation times were also the windows that had the maximum NCD scores. In fact, we could use calculation time as a proxy for the actual NCD score when identifying the seismic arrival times. This could be an interesting area for future research on compression metrics.

2.2.4. Optimizing the Movement of the Window

The final key issue required for applying NCD to waveforms was the movement of the window between calculations. At a minimum, the window could move by one sample for each new NCD calculation. However, most seismic data is sampled at 100Hz. If we choose a window width of 8 seconds, that is 800 samples. If we only move the window one bin at a time, that only changes one value out of 400 in each half of the window and produces a very similar NCD calculation. The NCD calculation is fairly time consuming (averaging 0.1-0.3 seconds per window for a window with 800 samples), so it is important to move the window in an efficient way. However, taking steps that are too large can make it more difficult to accurately locate the NCD peak value. It may be offset from the arrival time of the seismic signal in the data if the NCD calculations are too coarse.

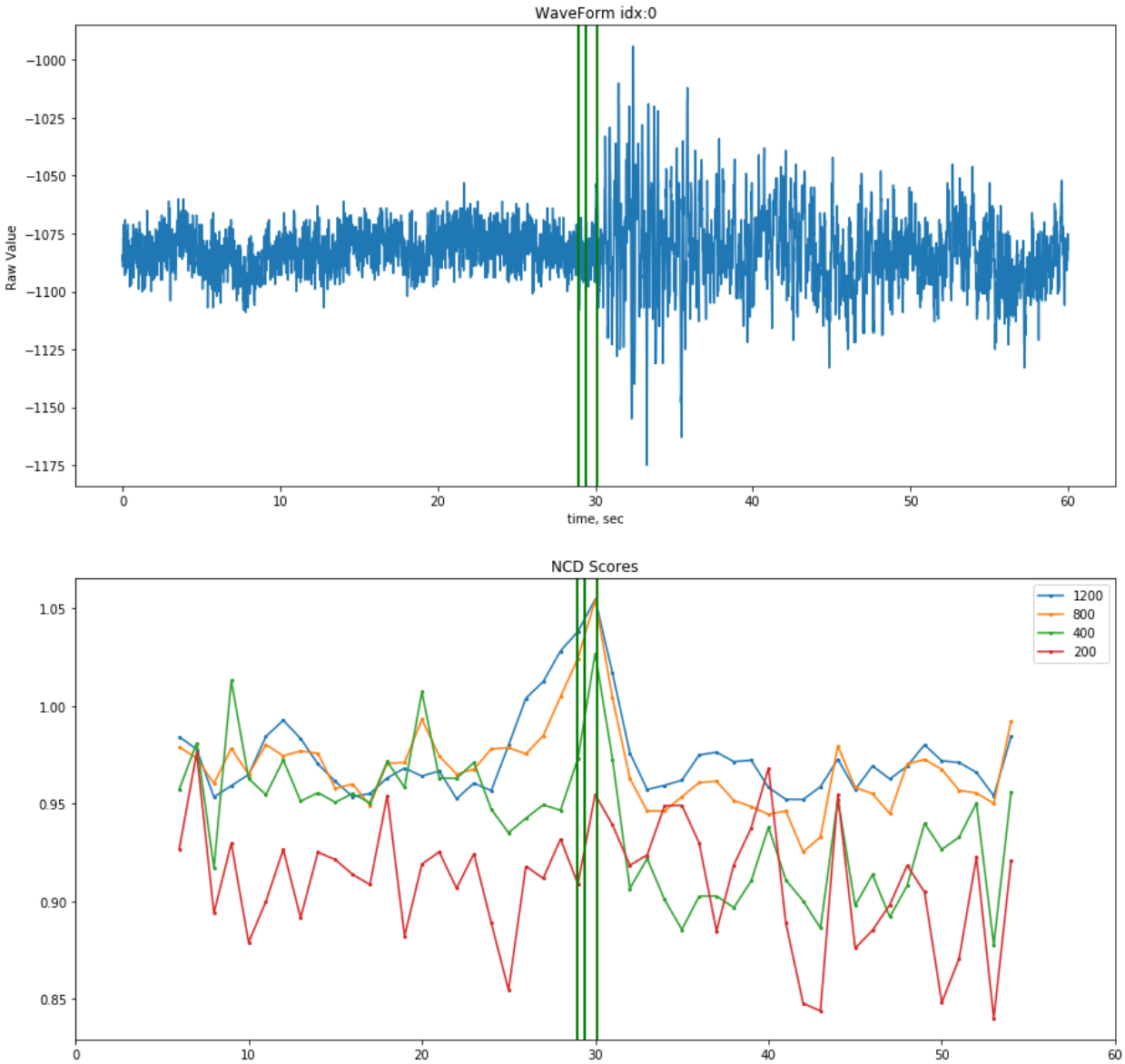

Figure 5. NCD scores calculated for the same seismic waveform using different window sizes. The vertical green lines represent the earliest, most likely, and latest possible arrival times for the seismic signal, as determined by an expert seismic analyst.

Based on analyses of a large dataset of seismic waveforms at differing window sizes and step sizes, we determined that moving the window so that 5% of the values in each half are updated provided a reasonable tradeoff between computational effort and the granularity of the analysis. For a window width of 800 samples, this would entail sliding the window by 40 samples between calculations. For seismic data sampled at 100 Hz, this produces an NCD value every 0.4 seconds. We also experimented with using a coarse-to-fine approach, where we made one pass through the waveform with large steps between windows, then computed a more fine-grained analysis with smaller steps around the time of the peak NCD value. However, our subsequent shift to using SLID instead of NCD (described in more detail below) made this coarse-to-fine approach unnecessary.

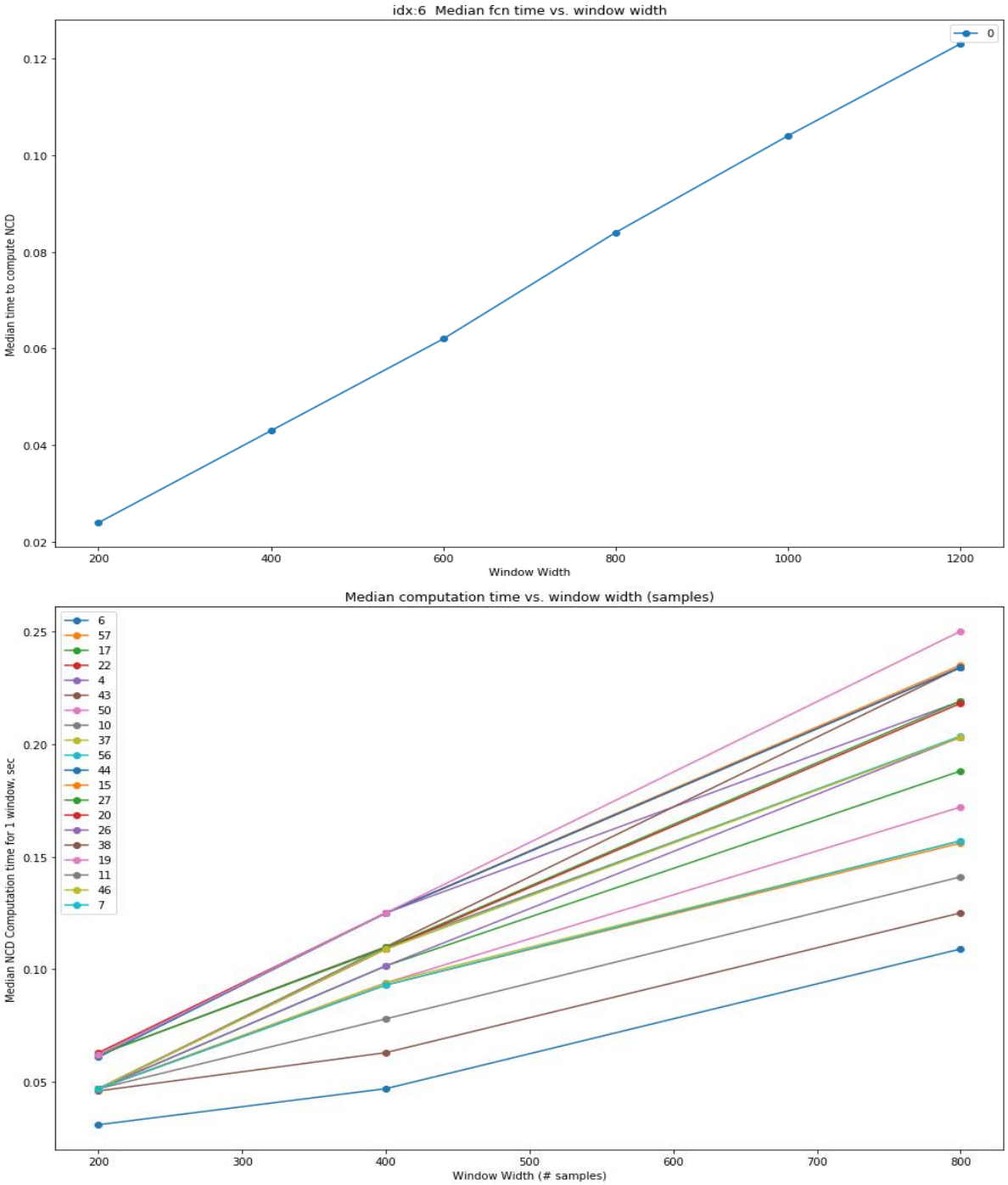

Figure 6. The median time to compute NCD for window widths of 200-1200 samples for one waveform (top) and for multiple waveforms from the same seismic event, recorded from different stations (bottom). This plot shows the signals from the 20 stations closest to the seismic event.

2.3. Using NCD to Identify Signal Arrival Times in Seismic Data

Having developed a method for applying NCD to seismic waveforms, our next step was to test how well it could identify signal arrival times across a diverse dataset. The accuracy of NCD was assessed

by comparing its predictions to manual "picks" made by a subject matter expert (SME). We used a dataset that was developed by Ronald (Chip) Brogan, a seismic analyst with over 27 years of experience with waveform data analysis in support of nuclear treaty monitoring. The dataset contained waveforms from 16 seismic events, which included earthquakes, mining activity, and nuclear tests. For each seismic event, waveforms showing the arrival of a seismic signal generated by the event were selected from two or more seismic stations. A total of 100 event/station combinations were selected, using data from 43 seismic stations. The majority of those stations are part of the University of Utah network[5]. The stations and a few of the events used in the dataset are shown in Figure 7. A table showing the combinations of seismic events and stations is shown in Appendix A.

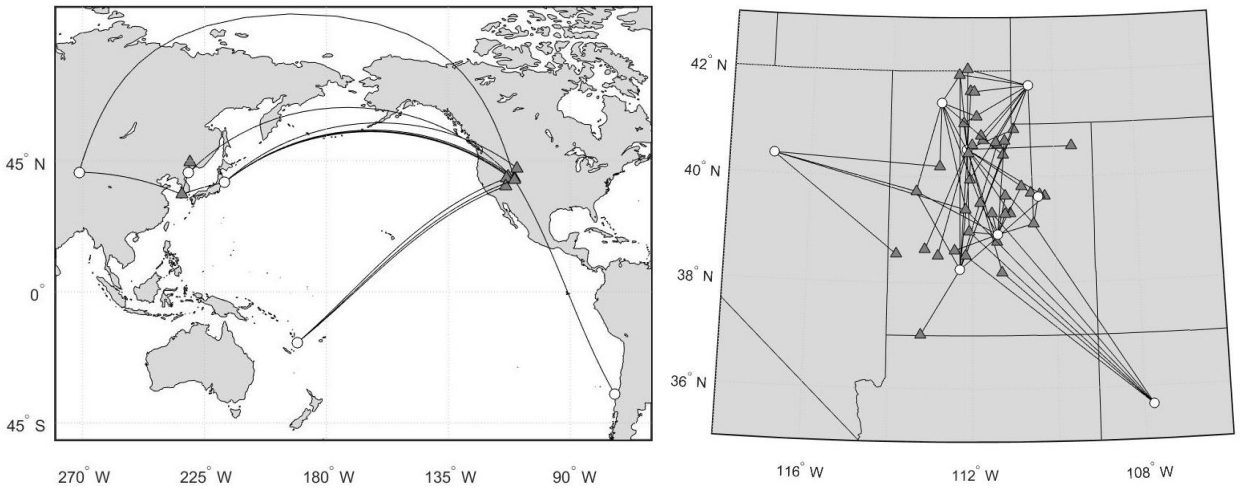

Figure 7. A world map (left) and a more detailed map of the area around Utah in the United States (right). The seismic stations used in this dataset are shown as gray triangles. A subset of the seismic events is shown as white circles, with lines drawn to indicate how the vibrations caused by each event travel to various stations.

For each event/station combination, the waveform showing the signal arrival could be filtered in various ways to make the signal more or less difficult to discern. The filters commonly used for seismic analysis are as follows:

1) Unfiltered
2) 0.8 Hz high-pass filter
3) 1-3 Hz bandpass filter
4) 2-4 Hz bandpass filer
5) 3-6 Hz bandpass filter
6) 4-8 Hz bandpass filter

[5] https://quake.utah.edu/monitoring-research/station-map The seismic stations included in this dataset were CHS, CTU, CWU, DAU, DBD, DCM, DUG, ELU, EMU, FLU, FSU, HLJ, HONU, HTU, ICU, IMU, LCMT, LEVU, LHUT, MLI, MMU, MSU, NAIU, NLU, NMU, OWUT, PNSU, PSUT, PTU, RBU, RCJ, RDMU, ROA, SNO, SRU, TCRU, TMU, WBC, WCU and WVUT from the University of Utah network, plus the stations MDJ and INCN from the Global Seismograph Network (https://earthquake.usgs.gov/monitoring/operations/network.php?virtual_network=GSN) and station PD31 from the International Miscellaneous Stations Network (https://www.fdsn.org/networks/detail/IM/).

Most, but not all, of the event/station combinations had waveforms available for all six filter bands, giving us a total of 536 waveforms in the dataset. For each of these waveforms, the SME made picks indicating the earliest, most likely, and latest possible signal arrival times. The picks were made independently for each filter type.

NCD was calculated for all of the waveforms in the dataset using a window size of 8 seconds. The window was shifted by 0.1 second between windows to account for the fact that some of the waveforms in the dataset had different sampling rates than others. The majority of the waveforms were sampled at 100 Hz, so the 0.1 second shift represented 10 samples. The NCD scores were normalized to a range of 0-1 for each waveform.

As a first pass, we determined how many waveforms contained a well-defined NCD peak by manually inspecting the NCD plots. A total of 423 of the waveforms (78.9%) had a clear NCD peak. The remaining waveforms fell into two groups: those that produced noisy NCD plots with many peaks of comparable magnitude, and those where the NCD value ramped up over time. Sometimes there was a peak that stood out within this slope, but other times there was not. Note that these ramping NCD peaks often corresponded to seismic signals that ramped up over time, making it difficult to determine the precise transition from pre- to post-event arrival.

To assess the performance of NCD quantitatively, the time at which the peak NCD value occurred was compared to the pick made by AIC, the current state-of-the-art algorithm for automatic event detections. Note that AIC picks are based on the minimum AIC value, while NCD picks are based on the maximum NCD value. NCD was also compared to the SME's "best" pick and the range of possible arrival times, as indicated by the SME's earliest and latest picks. We made these comparisons in two ways. First, we calculated the difference between the peak NCD score and the metric of interest for all 536 waveforms. Next, because some of the filters are better than others, we selected the filter band for which NCD performed best for each of the 100 events in the dataset. Then we compared the best-case scenario for each event to AIC and the expert picks for that same waveform.

2.3.1. Comparing NCD to AIC

When selecting the best filter for each of the 100 events, the difference between NCD and AIC was quite small. The results are shown in Figure 8. For 88 events, the difference between the predictions of the seismic signal arrival time using the two approaches was less than or equal to 0.1 seconds. The difference was less than 0.2 seconds for all but six of the events.

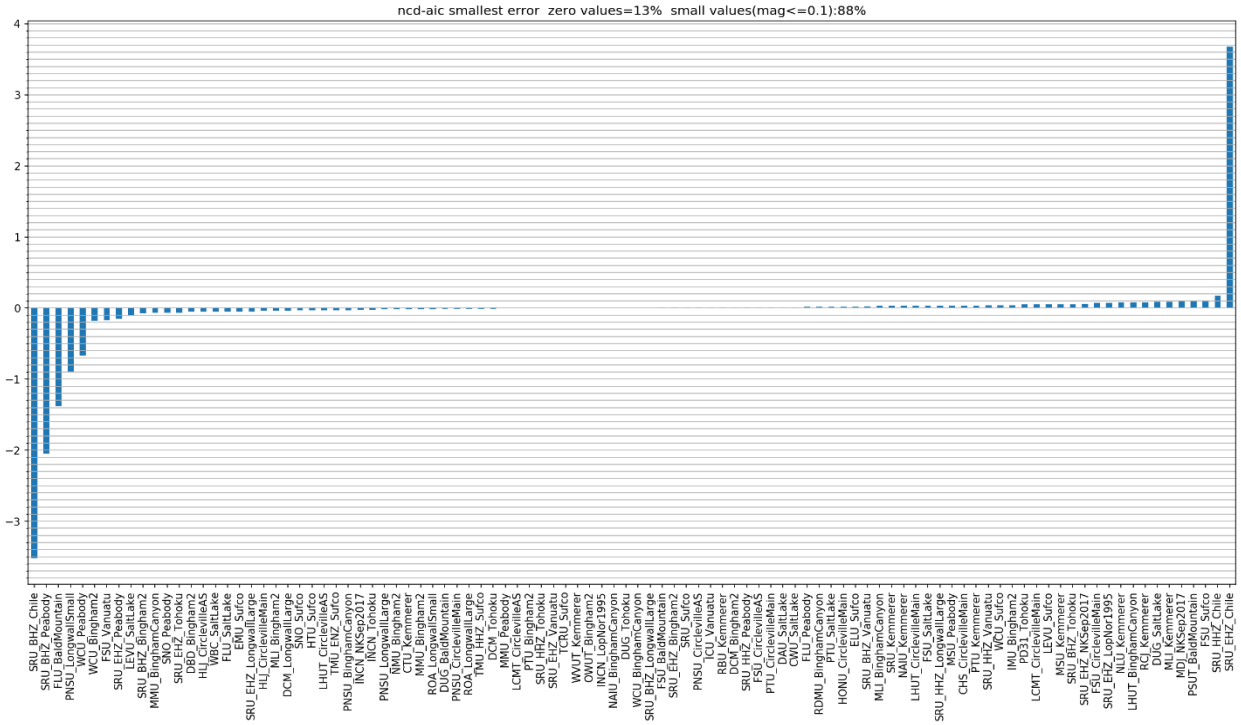

Figure 8. The smallest NCD-AIC difference for each of the 100 events in the dataset.

For the six events that had large differences between NCD and AIC, only one had data available for all six filters, so there was less to work with when choosing the best filter. The waveforms from these six events were very noisy and produced noisy results for both AIC and NCD. An example is shown in Figure 9. The NCD plot is shown in blue with a dashed blue line marking the maximum NCD score. The AIC plot it shown in orange with a line marking the minimum AIC score.

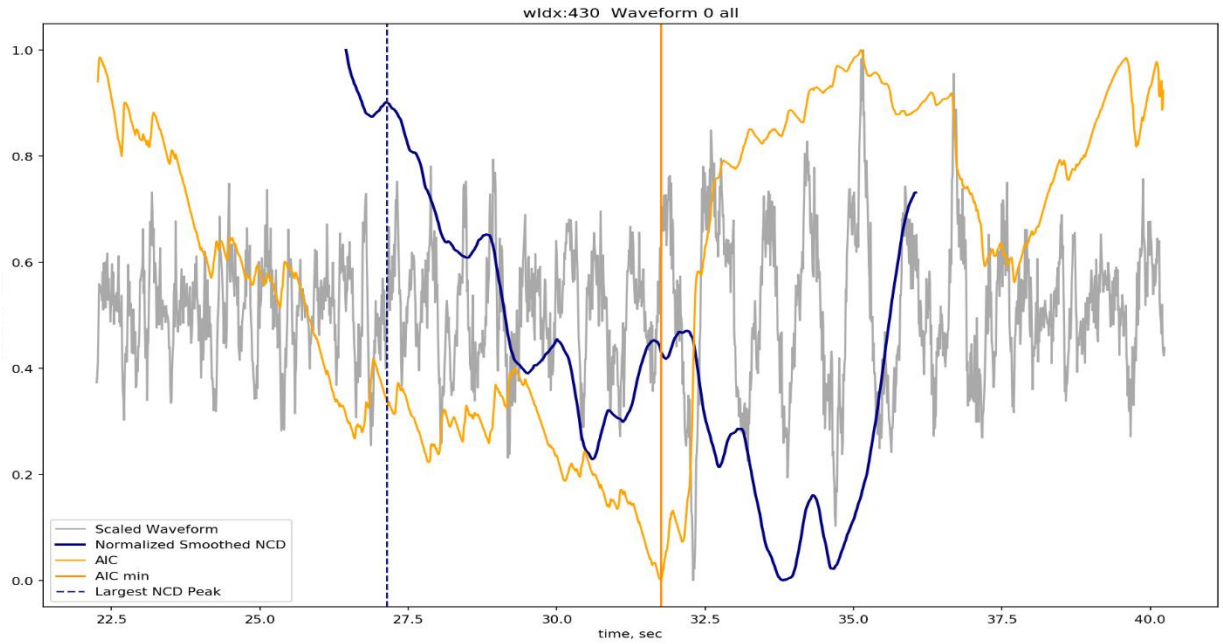

Figure 9. This example shows one of the largest NCD-AIC differences. Note that the waveform is very noisy, so the AIC and NCD plots are both noisy as well.

The vast majority of the events showed little to no time difference between the maximum NCD value and the minimum AIC value. Figure 10 shows a representative example. In cases where the signal arrival time is clear, both algorithms tend to have a sharp peak occurring at the same time.

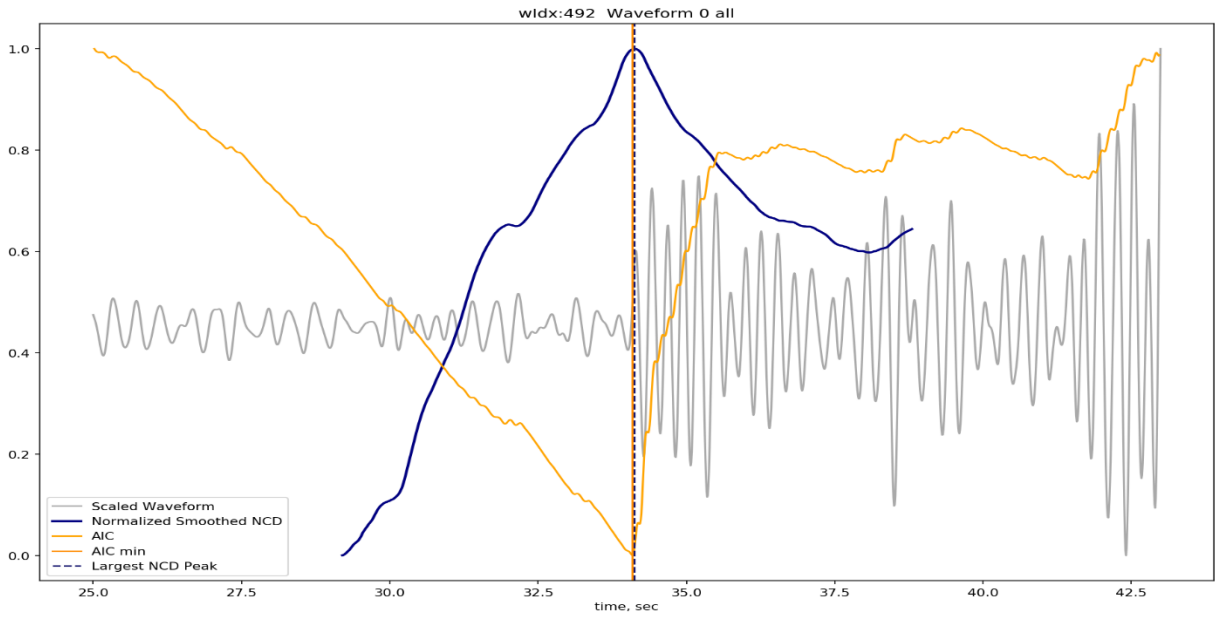

Figure 10. An example with no time difference between the NCD and AIC peaks.

2.3.2. Comparing NCD to the Best Expert Pick

Figure 11 shows the differences between the expert analyst's pick for the best (most likely) arrival time and the NCD peak for each event, using the filter for which that difference was smallest. These differences were somewhat larger than the differences we observed between NCD and AIC. There was also a bias where more of the events had a positive difference. This indicates that the NCD peak often occurred slightly later than the expert's pick for the best arrival time. However, the difference was still less than 1 second for all but three of the events and was less than 0.1 seconds for 49 of the 100 events.

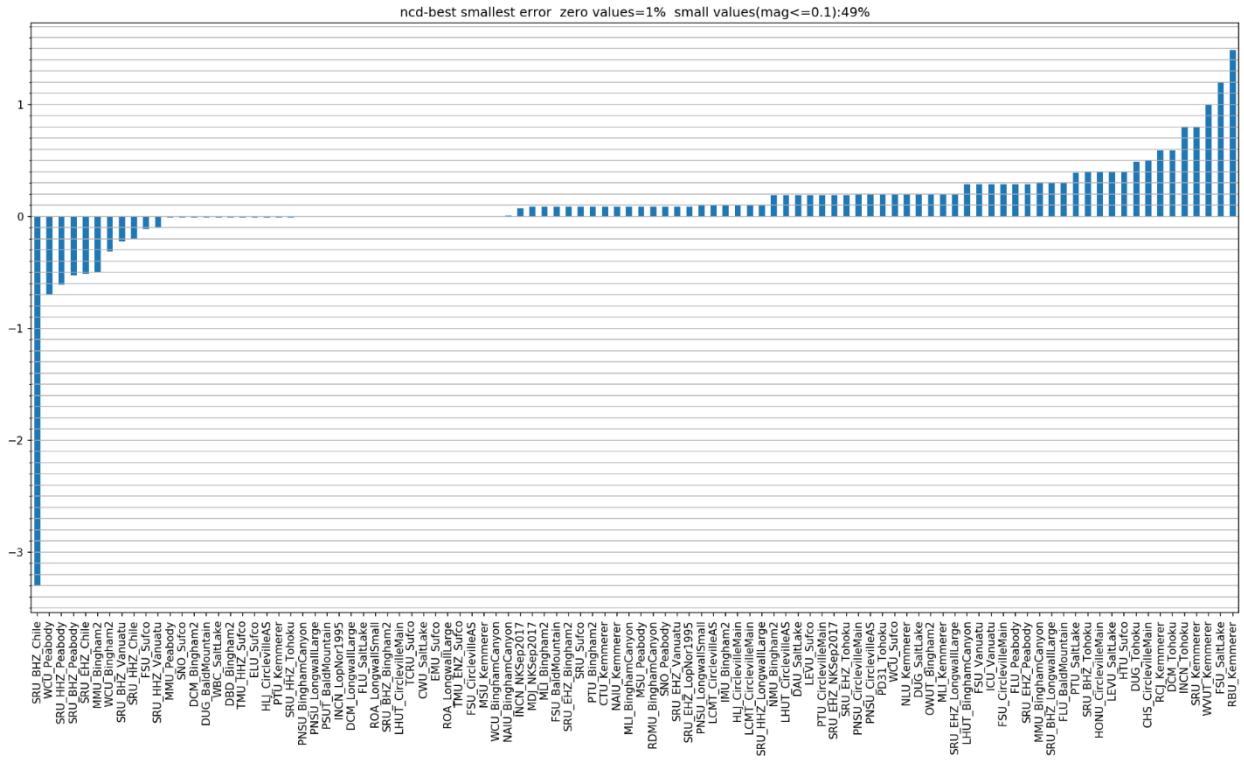

Figure 11. The smallest NCD-Best Pick difference for each of the 100 events in the dataset.

Figure 12 shows an example where the NCD peak aligns closely with the analyst's best pick, which is indicated with a green line. Figure 12 shows an example of one of the largest differences. In this case, the expert's pick is much earlier than the NCD peak. The NCD peak aligns with the point at which the amplitude of the signal begins to change. However, the SME picked an earlier time point as the most likely arrival time. The SME's process for making picks is to identify the amplitude change in the signal and then "look to the left" for a change in frequency. This strategy, plus knowledge that is external to the signal itself (such as knowledge about common patterns at particular seismic stations), led the SME to make a pick that is different than what an algorithm such as NCD can determine from the signal alone. However, note that NCD was somewhat sensitive to the change identified by the SME, as indicated by the local maximum that is visible in the NCD plot in Figure 13.

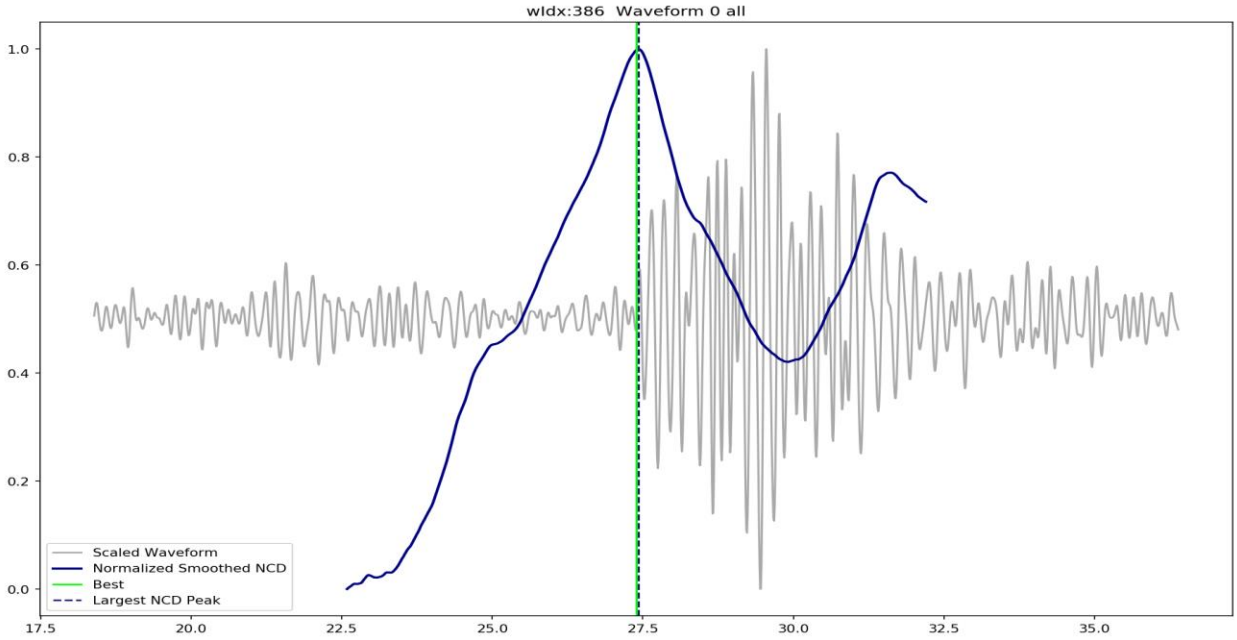

Figure 12. An example with a close alignment between the NCD peak and the analyst's best estimate of the signal arrival time.

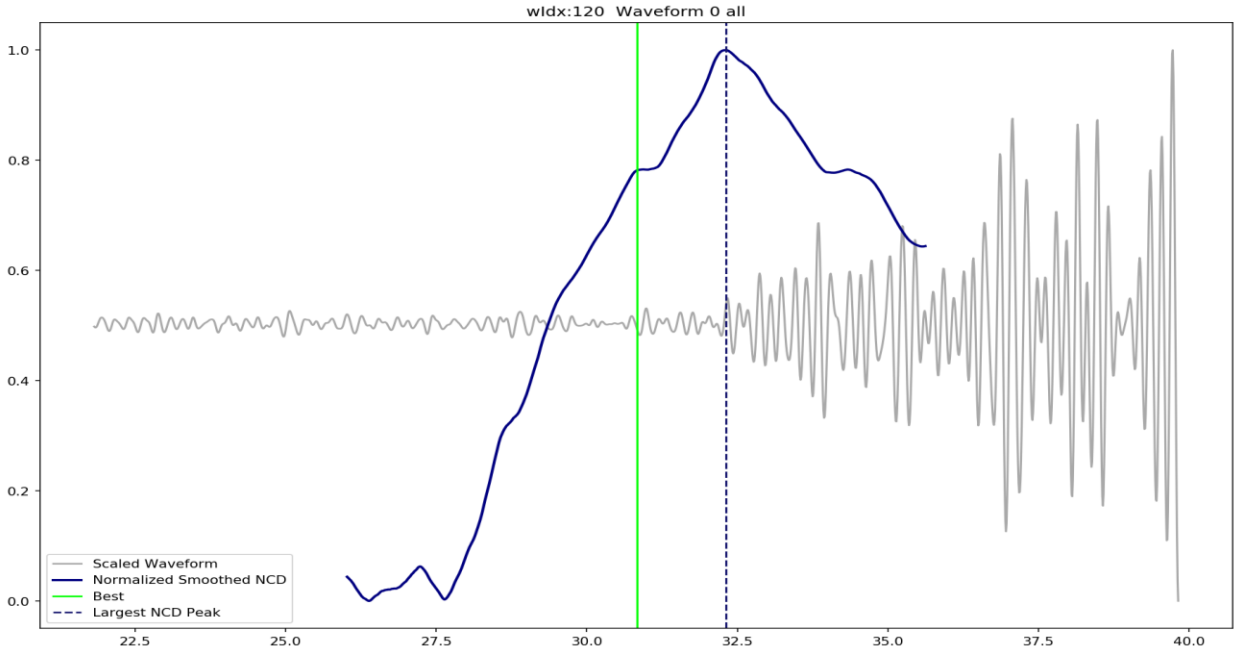

Figure 13. An example where the NCD peak is late relative to the analyst's best pick.

2.3.3. Comparing NCD to the Range of Possible Arrival Times

Figure 14 shows the differences between the NCD peak and the expert analyst's arrival band. The arrival band represents the analyst's assessment of the earliest possible and latest possible arrival times. 75% of the NCD peaks fell within the arrival band and 87% had an error of less than 0.1

seconds. Once again, we observed a slight bias for the NCD peak to occur slightly later than the expert picks, but this difference was quite small.

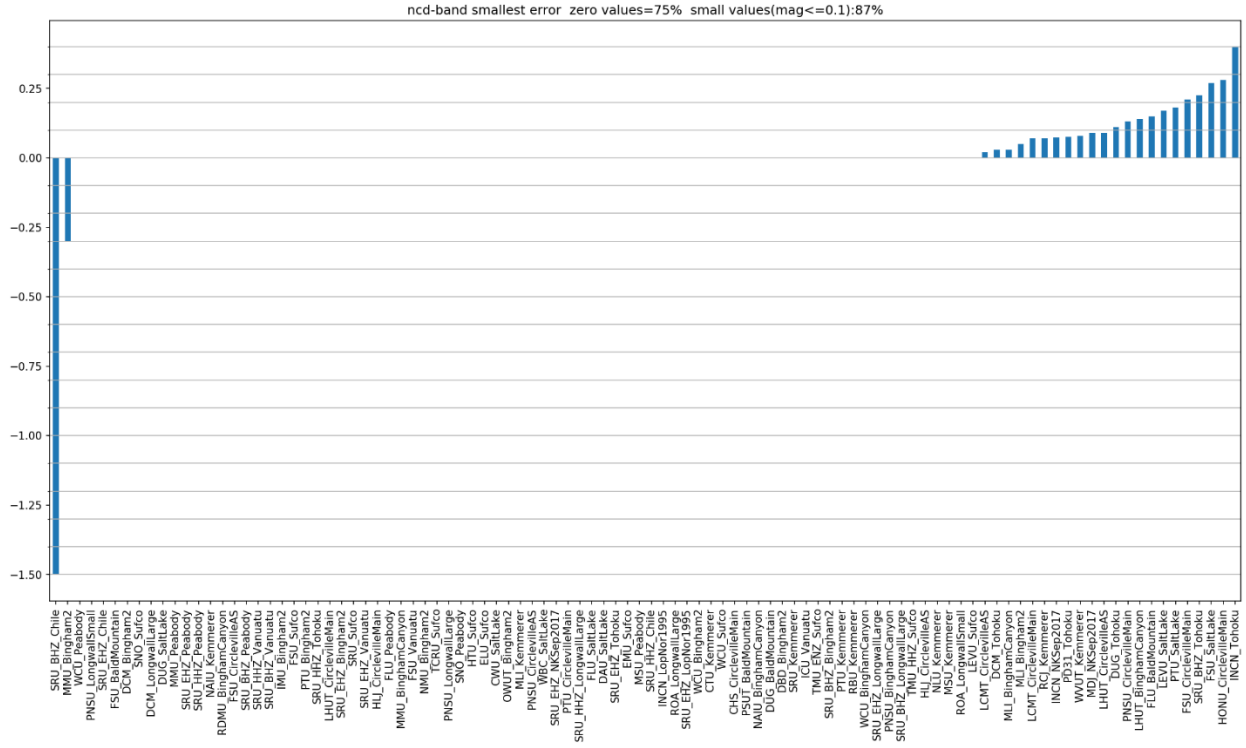

Figure 14. The smallest NCD-Arrival Band difference for each of the 100 events in the dataset.

Figure 15 shows an example where the NCD peak falls just inside of the arrival band, which is represented by the shaded green region. Note that the NCD peak is slightly later than the analyst's pick for the best arrival time, which is represented by the lime green line. However, it still falls before the analyst's estimate for the latest possible arrival time. Figure 16 shows one of the largest errors. In this case, the NCD peak occurs much later than the arrival band. Note, however, that there is a local maximum in the NCD plot that falls within the arrival band.

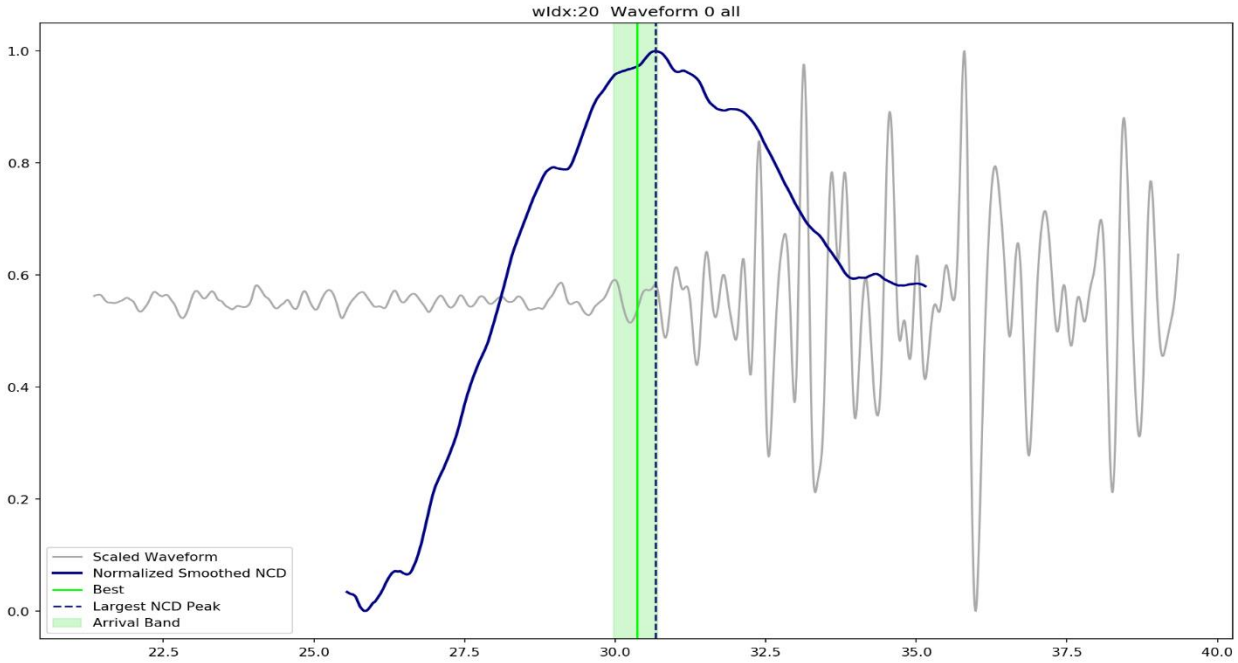

Figure 15. An example where the NCD peak falls just inside the arrival band.

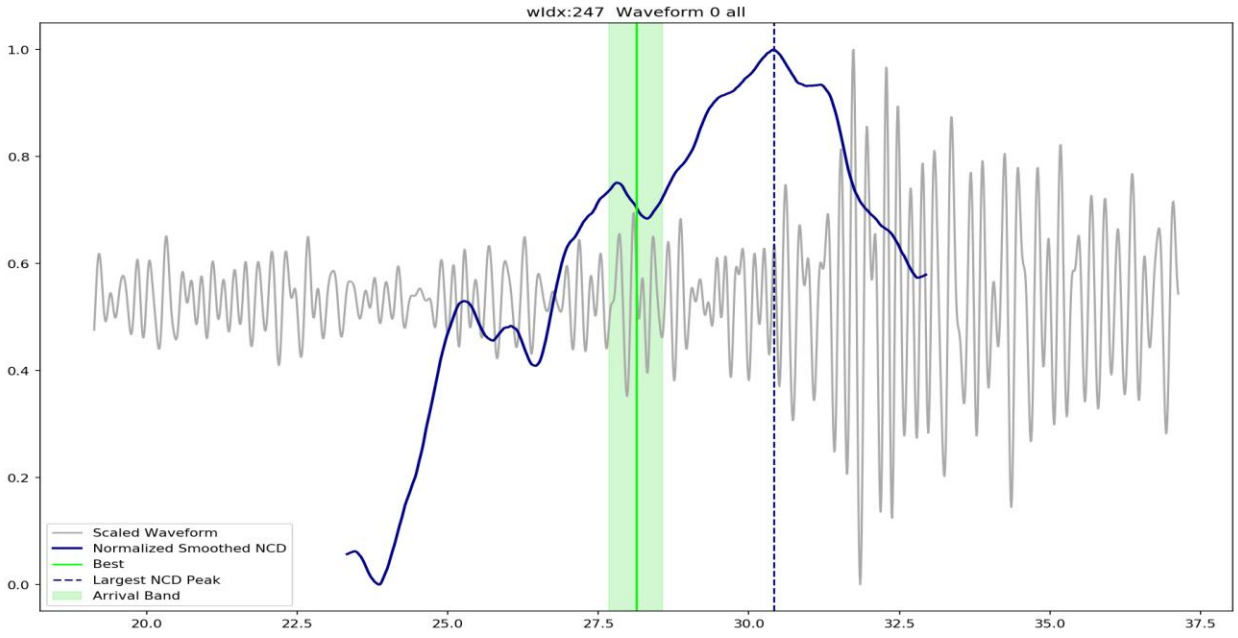

Figure 16. An example where the NCD peak is considerably later than the arrival band.

2.3.4. Summary of Findings

Overall, our results indicate that the maximum NCD score for a seismic waveform provides a good estimate of the arrival time of the seismic signal. The NCD results corresponded closely to the AIC results. The NCD peak often fell slightly later than the expert analyst's picks. There are two possible reasons for this slight bias. One is the moving window implementation of NCD, which could cause

the peak to fall late relative to the signal arrival if the steps being used between windows are relatively large. Another reason for this bias could be the fact that NCD, as implemented here, does not directly consider changes in the frequency content of the waveforms. It is primarily sensitive to changes in amplitude. The expert analyst explained that he looks for changes in amplitude in the signal, then works backwards looking for changes in frequency. His goal is to find the earliest possible indicator of the event arrival. This strategy is likely to lead to earlier picks than NCD because it is deliberately searching for the earliest possible change, moving backwards in time from the transition point that produces the highest NCD scores. In future work, it would be useful to explore ways to incorporate frequency information into NCD and other compression metrics.

In many cases where the maximum NCD peak was late relative to the expert picks, we observed that there was a secondary peak in the NCD plot that corresponded more closely to the best pick. Examples of this pattern can be seen in Figures 13 and 16 as well as in Figure 17, below. In Figure 17, the earlier NCD peak is marked with a light blue line and falls much closer to the expert analyst's pick.

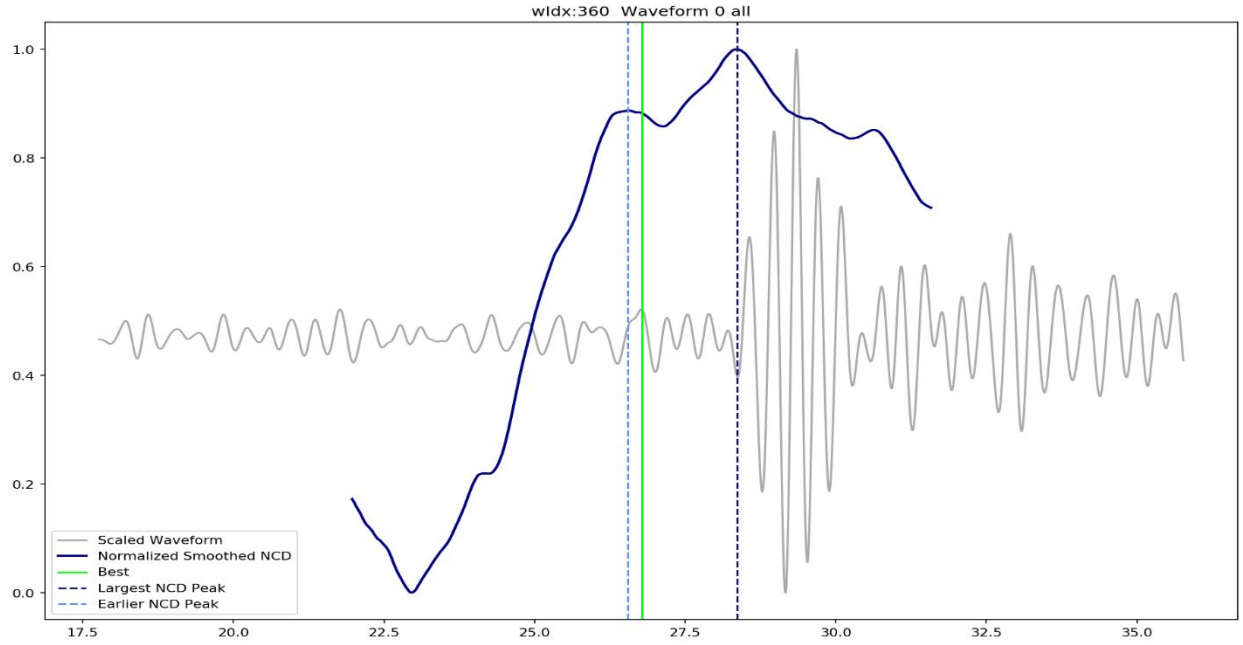

Figure 17. An example where the earlier NCD peak is closer to the expert's pick.

Based on this pattern, we developed a method for detecting and marking secondary peaks in the NCD plots. We focused on secondary peaks that occur earlier than the primary peak, due to the tendency for the primary peak to occur slightly later than the expert picks. For the purposes of our research, we marked earlier peaks if there was a local maximum that was at least 80% as tall as the primary peak occurring within the 5 seconds of the primary peak. The impact of including these earlier peaks in the analysis is outlined in the following section.

2.3.5. Is NCD or AIC Closer to the Expert Analyst's Picks?

Although the NCD and AIC results were typically very similar, it is useful to know which algorithm produces results that are most similar to expert judgments. To assess this, we calculated the difference between AIC's pick (the time of the minimum AIC value) and the expert analyst's best

pick for all 536 waveforms. Then we did the same for NCD, defining NCD's pick as the time of the maximum NCD value. When there was an earlier NCD peak that met the criteria outlined in the previous section, we compared the times for both the primary peak and the earlier peak to the expert's best pick. Of the 536 waveforms, there were 114 that had an earlier peak that met our criteria. For 61 of those waveforms, the earlier NCD peak was closer to the expert's best pick than the primary peak. For each waveform, the algorithm that produced a pick that was closer to the expert analyst's best pick was deemed the winner.

Across all 536 waveforms, NCD won 365 times while AIC won 170 times. For one waveform, the differences were exactly equal. In some cases, both algorithms produce picks that are less than 0.1 seconds from the expert's pick. Since winning is not particularly meaningful when both results are extremely close to the expert's pick, we opted to exclude those 74 waveforms from the analysis. When those small errors were excluded, NCD won 324 times and AIC won 135 times.

This analysis showed that NCD can provide better automatic picks than AIC, in the sense that they are often closer to the picks made by an expert analyst. These results also indicate that the shape of the NCD plot itself can provide useful information. For example, taking the shape of the plot into account allowed us to identify earlier peaks that may be closer to the arrival time of the signal. After analyzing the NCD results for this dataset, we hypothesized that showing a plot of compression distance information to analysts could help to inform their decision making. One way this information could be useful is for data triage. If NCD and AIC produce very similar picks, there should be no need for a human analyst to adjust the pick times. Agreement between these two algorithms indicates that the signal arrival time is very clear. Similarly, when NCD has a very sharp peak, that indicates that the information content of the waveform supports making a pick that corresponds to the peak of the NCD plot. Placing the pick elsewhere is not likely to improve the analysis. In contrast, when NCD does not produce a single peak, the shape of the NCD plot can help analysts to narrow in on the most likely signal arrival times. When there is more than one NCD peak, analysts may want to inspect both regions. The shape of the NCD plot could also be used to automatically select the filter in which the signal is clearest. This is another use case that could save seismic analysts time and effort.

2.4. Using NCD as a Signal Rectifier for Event-Level Analyses

Our work applying NCD to individual seismic waveforms showed that it is an effective method for identifying likely signal arrival times. Identifying arrival times across multiple seismic stations is a key step in building seismic events, so it may be possible to use NCD information for signal rectification when building seismic events.

To test this, we calculated NCD for 50 waveforms collected from 50 different seismic stations for the same event. When the plots were ordered by distance from the seismic event, the NCD plots showed the expected shift in time, with the NCD scores peaking later at stations that are farther from the seismic event. The NCD picks also corresponded closely to picks made by an expert analyst, as shown in Figure 18.

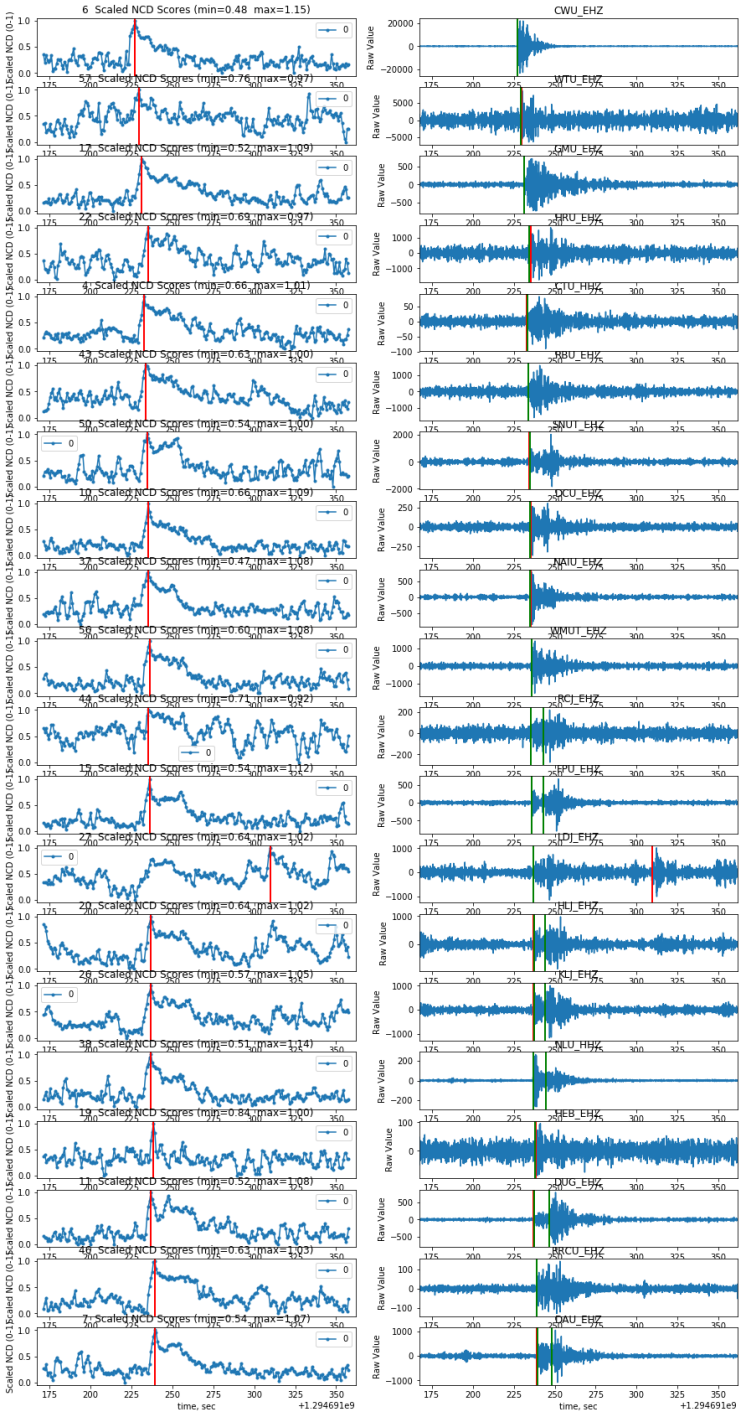

Figure 18. The NCD plots (left) and seismic waveforms (right) with stations ordered by distance. The maximum NCD value, marked with a red line, is shown in all plots. It typically corresponds closely to the analyst's picks, shown with green lines.

2.5. Summary of NCD Work

Our primary goal for this project was to determine whether compression metrics such as NCD could be applied to seismic waveform data and whether this would produce information that could be helpful to seismic analysts. We successfully developed a moving window approach for calculating

NCD metrics on seismic data. By testing the impact of different discretization methods, window sizes, and step sizes, we identified parameters that are well-suited to the characteristics of seismic data. In comparing the results of the NCD calculations to AIC and expert picks, we showed that NCD typically produces a peak that is close in time to expert picks, and it is often closer to the expert judgement than picks based on AIC. NCD can also be used as a signal rectifier, which can help analysts determine if a given seismic signal corresponds well to a larger event.

One drawback of NCD is that it frequently peaks slightly late relative to a SME's determination of the signal arrival time. As discussed above, this slight delay is most likely due to NCD's sensitivity to amplitude rather than the frequency content of the waveforms. Since NCD does not explicitly model the frequency content of the waveforms, there is a slight mismatch between the NCD peak and the strategy that was used by the SME, which emphasizes frequency changes occurring slightly before amplitude changes. It is important to note that in addition to looking at the amplitude and frequency of the waveforms, human analysts bring external knowledge to bear when making picks in seismic waveforms. Information that is not evident in the waveforms themselves, such as knowledge about the types of noise that occur at a particular seismic station or common patterns that have been observed from past seismic events, are not available to NCD or any other algorithms that must rely solely on the data at hand.

Despite this drawback, NCD provides several other advantages over existing methods for seismic analysis, such as AIC. By plotting the NCD values over time, we can identify waveforms where the automatic picks are well-supported, select the optimal filter for each waveform, and identify the segments of a seismic waveform that are most likely to contain the signal onset even when the data are very noisy. All of these cues could help analysts to triage data, focusing their efforts on the examples that require expert judgement and external knowledge, rather than spending time adjusting picks that are already well-supported by the information content of the waveforms themselves.

The primary remaining drawback of NCD is the time it takes to compute. It can take 0.1-0.3 seconds to compute NCD for each window. Although using parallel processing can speed this up, it would still be very time consuming to calculate NCD for the 8-hour segments of data that analysts typically work with. To some extent, this processing could be done offline, so that the NCD calculations are available when an analyst starts his or her shift. But ideally, we would like seismic analysts to be able to call up compression metrics in real time. To meet this goal, we began testing an alternate compression metric: sliding information distance (SLID). SLID is specialized for change point detection and can be computed much more rapidly than NCD. In the sections that follow, we will introduce SLID, discuss how it can be applied to seismic waveforms, and assess its performance relative to NCD.

3. APPLYING SLIDING INFORMATION DISTANCE TO SEISMIC WAVEFORMS

As discussed above, NCD takes a non-trivial amount of time to compute. The overhead of the full compression algorithm $C(\cdot)$ may render the NCD impractical for use in change point detection for real-time analysis systems. Since the NCD needs to be computed repeatedly as we progress the adjacent windows through the seismic waveform, and analysts are dealing with thousands of waveforms, it would be too time-consuming to calculate NCD on the fly.

Due to the computational overhead for NCD, we began to explore alternative compression metrics. It has been shown that the computational complexity of approximating the NID can be significantly reduced by operating directly on the underlying dictionaries constructed by the compression algorithms (Cerra & Datcu, 2012; Koga, Nakajima & Toda, 2016; Macedonas et al., 2008; Raff & Nicholas, 2017; Ting et al., 2019]. The *sliding information distance* (SLID) is a variant of these methods further specialized for change-point detection. We have previously shown that SLID efficiently, accurately, and robustly approximates the NID as it progresses, or *slides*, over adjacent windows in general data sequences (Ting et al., 2019). In this section, we demonstrate the application of SLID to seismic waveforms.

3.1. Calculating the Sliding Information Distance

Let k denote a position within the waveform $z = \{z_0, z_1, \dots\}$. We formulate the event detection problem by considering two adjacent windows of z, denoted by x_k and y_k, each of length $w \geq 1$:

$$\mathrm{z}_0 \dots \mathrm{z}_{\mathrm{k-w-1}}, \underbrace{\mathrm{z}_{\mathrm{k-w}} \dots z_{k-1}}_{x_k(w)} \underbrace{\mathrm{z}_{\mathrm{k}} \dots \mathrm{z}_{\mathrm{k+w-1}} \mathrm{z}_{\mathrm{k+w}}}_{y_k(w)} \dots$$

To obtain the SLID score of z at position k, each window $x_k(w)$, $y_k(w)$ is first converted to a dictionary of observed subsequences $X_k(w)$, $Y_k(w)$ using a method that relies on a simplified version of the Lempel-Ziv (LZ) algorithm (Ziv & Lempel, 1977; Ziv & Lempel, 1978). For notational conciseness, we drop the dependence on w and refer to X_k , Y_k as the *LZ sets* of windows x_k, y_k. The SLID score is then simply the Jaccard distance between two adjacent LZ sets and is defined by

$$\mathrm{S(k)} = \begin{cases} 1 - \dfrac{|X_\mathrm{k} \cap Y_\mathrm{k}|}{|X_\mathrm{k} \cup Y_\mathrm{k}|}, & X_\mathrm{k} \cup Y_\mathrm{k} \neq \emptyset \\ 1, & otherwise \end{cases}$$

SLID takes values in $[0,1]$. Algorithm details on the computation of SLID are presented in Ting and colleagues (2019).

Thus, SLID transforms a seismic waveform z into a signal $S(k)$ whose value at a position k corresponds to an approximate information distance between adjacent windows, each of size w. For the seismic application considered in this document, this resulting signal is postprocessed by applying a smoothing window of size u. Larger values in both w and u will produce smoother SLID signals. In general, windows large enough to average out the noise without averaging out the large-scale structure are desired.

3.2. Applying SLID to Seismic Waveforms

Once the smoothed SLID signal is calculated for a given waveform, it remains to identify seismic events. We assume seismic events are *peaks* within the SLID signal. Peaks within a signal are easy for a human to identify. However, it is not straightforward for an algorithm to identify peaks. We describe one approach for peak finding based on identification of *peak regions* and merging of adjacent peak regions. A schematic of the event detection workflow is shown in Figure 19.

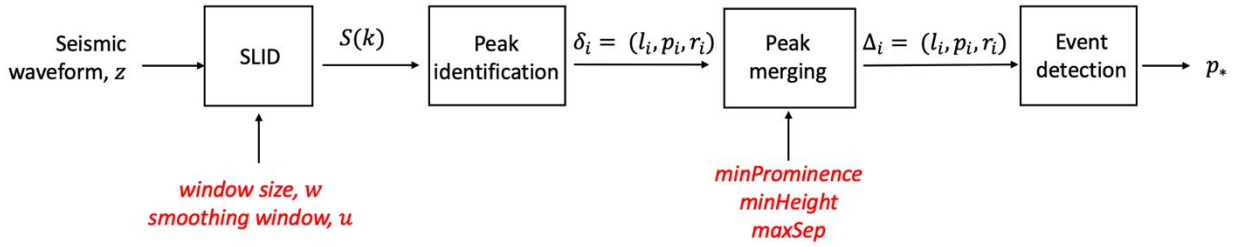

Figure 19. Flow chart for event detection of seismic waveforms.

3.2.1. Peak identification

To begin, we define peak regions as regions that are locally concave. Each region can be described by a list of indices describing the left, peak, and right positions within the region: $\delta_i = (l_i, p_i, r_i)$ where the indices satisfy $l_i < p_i < r_i$ and the SLID signal satisfies $S(l_i), S(r_i) < S(p_i)$. Because the seismic signal is noisy (even after smoothing) there will most likely be many candidate peak regions. Therefore, a key component of the algorithm is merging of peak regions.

3.2.2. Peak Merging

The main steps to merging of candidate peak regions can be described as follows:

1. For each candidate peak region δ_i, compute the left and right prominences, that is, the change in magnitude between the value of the peak $S(p_i)$ and the bounds $S(l_i), S(r_i)$.
2. For regions whose left and right prominences are less than some tolerance parameter `minProminence`, search for adjacent peak regions, defined as regions whose indices are within `maxSep` and merge adjacent peak regions until the new left and right prominences are larger than `minProminence.`
3. Finally, filter out merged peak regions whose peak values are below a threshold denoted by `minHeight.`

The above algorithm produces a list of *merged* peak regions $\Delta_i = (l_i, p_i, r_i)$ consistent with the specified values for parameters `minProminence`, `minHeight`, and `maxSep`. If n_Δ is the number of merged peak regions and n_δ is the number of candidate peak regions, then $n_\Delta \leq n_\delta$.

3.2.3. Event prediction

Given the collection of n_Δ merged peak regions, the next step is to identify the seismic event of interest. For general change point detection applications, it is possible to take all n_Δ merged peak regions as event regions. However, for the seismic event detection application, we have external

knowledge that there is a single event per waveform. Therefore, our prediction of the event is simply the time of the largest peak:

$$p_* = \underset{i=1,\ldots,n_\Delta}{\operatorname{argmax}} S(p_i)$$

3.3. Uncertainty Quantification with SLID

The prediction p_* of the seismic event onset time depends on the choice of the five key parameters, which are shown in red in Figure 19. Without prior knowledge on the optimal parameter set, it is difficult to determine the confidence of our prediction for a given choice of parameters.
One approach to assess the confidence in our prediction p_* is based on uncertainty quantification (UQ). Specifically, if we assume the five parameters are mutually independent uniform random with the following ranges:

- `5 < windowSize < 20`
- `0 < smoothingWindow < 5`
- `0.005 < minProminence < 0.2`
- `0.5 < minHeight < 0.8`
- `0 < maxSep < 2`

Our approach is to draw $n = 100$ random samples from these parameters and run the SLID event detection algorithm for each of the 100 random samples, producing 100 predictions for a single seismic waveform. The resulting probability density function (pdf) of p_* is fitted with a kernel density estimator (KDE) with a Gaussian kernel. The standard deviation of the fitted distribution is inversely proportional to confidence.

Figure 20-22 illustrate this approach on the SLID seismic event predictions. In addition to the distribution of p_* (red), we also plot the distribution of all merged peaks, p_i , for $i = 1, \ldots, n_\Delta$ (yellow). For reference, the expert picks (best and range) are also shown (blue). It can be seen that the standard deviations of the distributions match our intuition for predictions with high, medium, and low confidence.

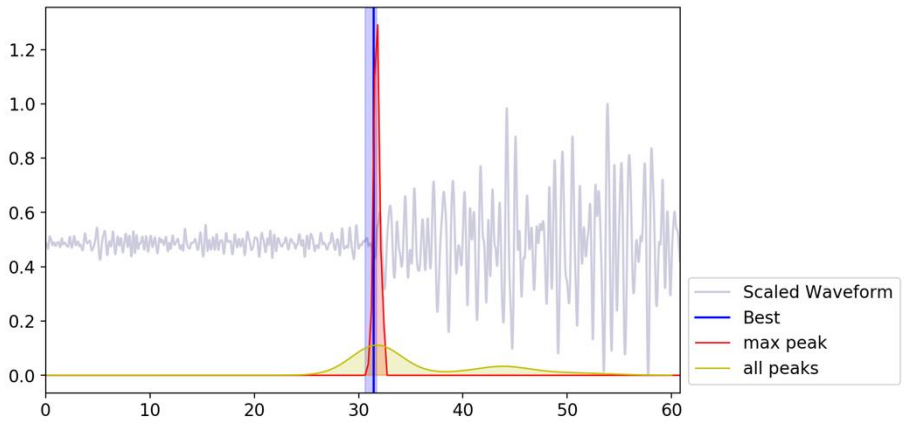

Figure 20. Exemplar result for a seismic event prediction with high confidence.

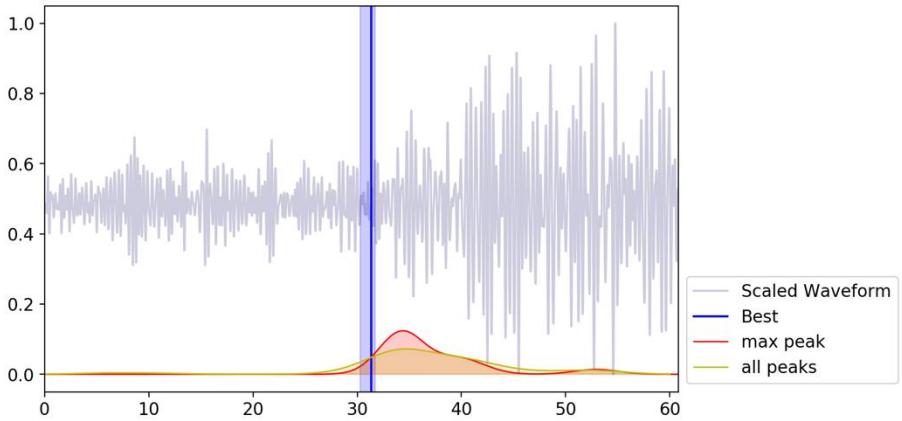

Figure 21. Exemplar result for a seismic event prediction with medium confidence.

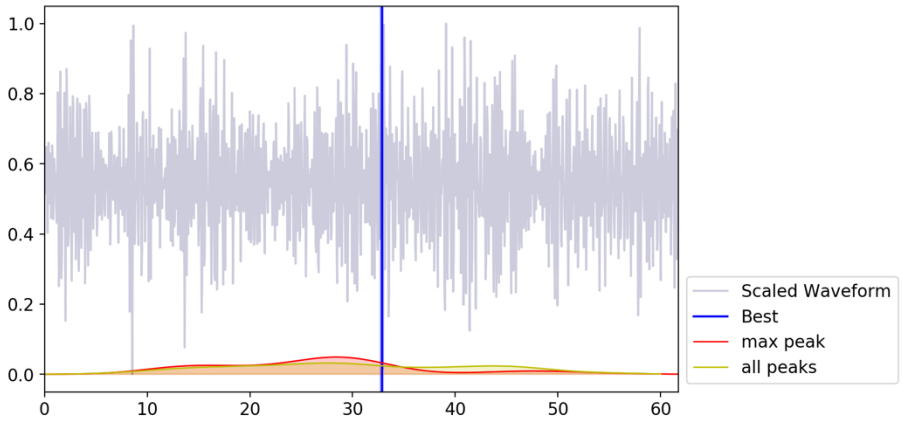

Figure 22. Exemplar result for a seismic event prediction with low confidence.

A practical use case for the confidence values is to allow the analyst to quickly sift through waveforms where the algorithm has high confidence, and to more carefully analyze waveforms where the algorithm has low confidence. In order for this use case to hold, the algorithm confidence must correlate with algorithm prediction performance. Therefore, we next look at the correlation between algorithm confidence and performance.

To quantify the algorithm performance, we compare μ, the mean in the distribution of the algorithm prediction time, with t_{SME}, the SME prediction time. Specifically, since performance should be inversely related to the difference between the mean algorithm prediction and the SME prediction, we look at

$$\text{performance} = |t_{SME} - \mu|^{-1}.$$

To quantify the algorithm confidence, we use the inverse of the standard deviation σ in the distribution of the algorithm prediction time:

$$\text{confidence} = \sigma^{-1}.$$

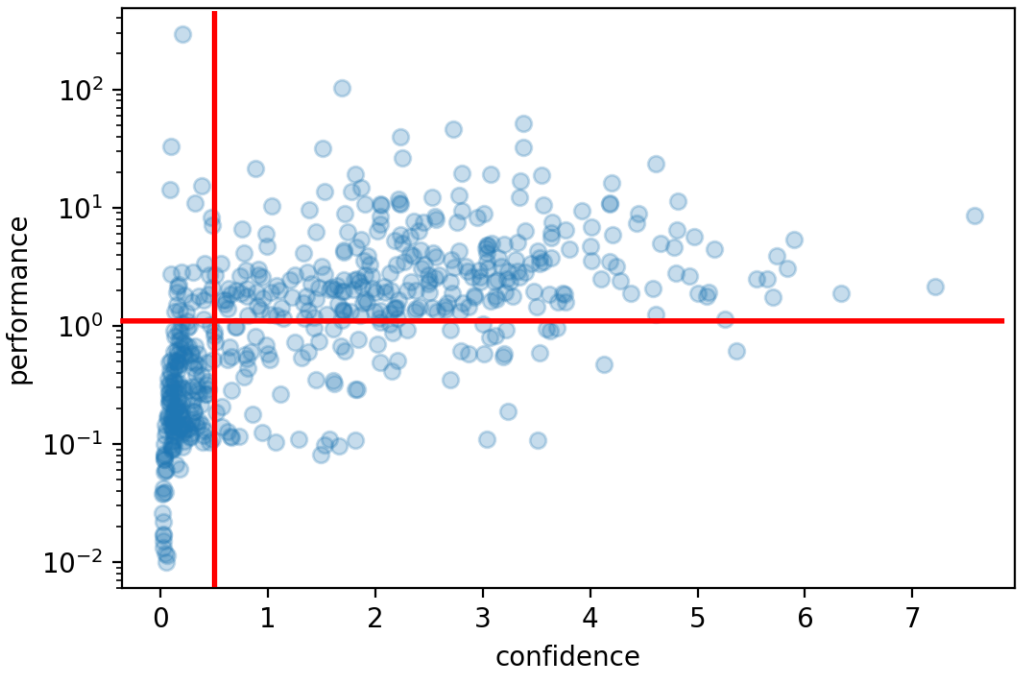

Figure 23. Relationship between algorithm performance and algorithm confidence.

The relationship between algorithm confidence and algorithm performance is shown in Figure 23. It can be seen that, in general, higher algorithm confidence tends to yield better algorithm performance. The highest density of low performing seismic waveforms (off by larger than 1 second from the analyst pick) also has a confidence score less than ~0.5 (bottom left corner in the plot). These results suggest that it is a promising approach to specify a threshold in the computed algorithm confidence score for a given seismic waveform (0.5 for this dataset), below which it is necessary to require an expert to make the event pick.

Finally, if we define waveforms where $|t_{SME} - \mu| \leq 1$ second to be a true positive, we can also assess algorithm performance using a precision recall (PR) curve. Intuitively, precision refers to the percentage of results that are correct; recall refers to the percentage of correct results returned. Ideally, one would like both high precision and high recall. Figure 24 shows the PR curve, obtained by ranking waveforms according to their confidence score (solid) compared with a random ranking of waveforms (dashed). It can be seen that returning waveforms to an analyst sorted by confidence score yields a significantly better PR curve (average precision (AP) = 0.87) than returning the same waveforms at random (AP = 0.56). Even for low recall values (~0.1), a random ordering of waveforms yields a precision of approximately 0.6, whereas the waveforms with the highest confidence yields a precision higher than 0.95. In fact, using the confidence score to return waveforms to the analyst, the precision stays higher than 0.8 until high values of recall. Note that at full recall, where we have returned all the true positives, we have a precision of approximately 0.6, indicating that approximately 60% of all waveforms are considered true positives.

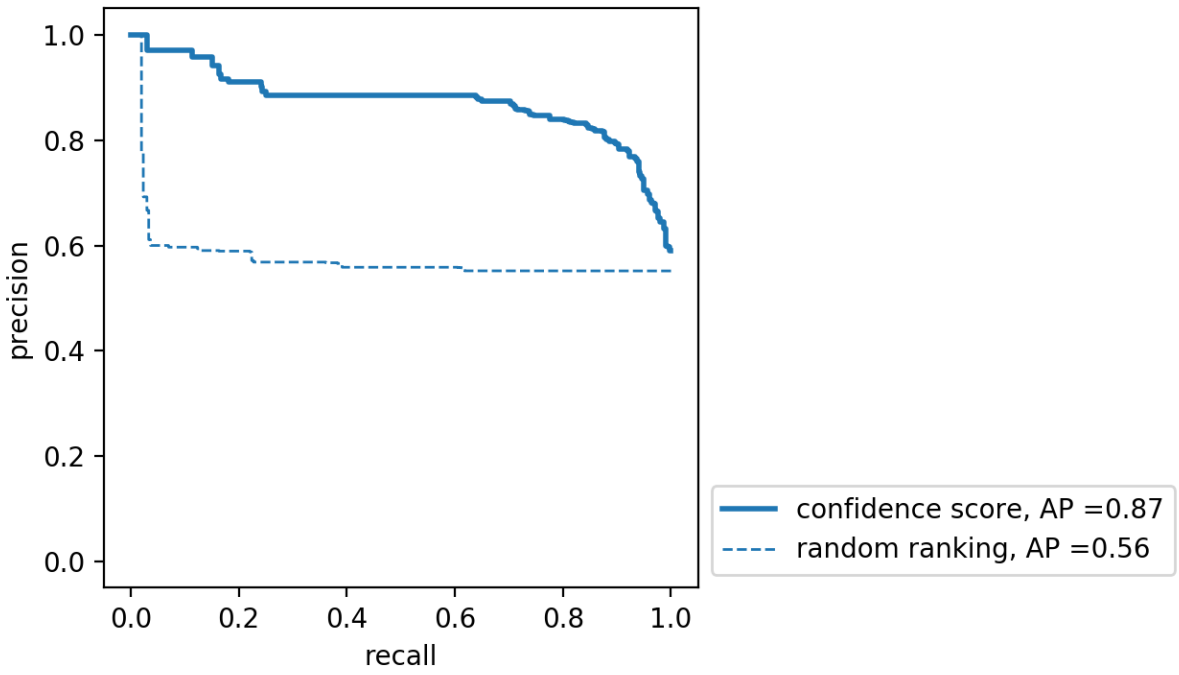

Figure 24. Comparison of precision-recall (PR) curve using the confidence score (solid) and a random ranking of waveforms.

3.4. Assessing SLID Performance for Seismic Data

Our initial work with NCD demonstrated that the maximum NCD value was typically close in time to the minimum AIC value and the range of possible arrival times picked by an expert analyst. SLID has major advantages over NCD from a computational perspective. It can be computed fast enough to be available on demand in real-time analysis settings, and it can be computed multiple times to provide UQ information for seismic waveforms. However, we wanted to ensure that it performed as well as NCD in terms of producing peak values that align closely with AIC and expert picks. To assess this, we compared the differences between the SLID peaks and the NCD peaks.

3.4.1. Comparing SLID to NCD

As a first step in this assessment, we calculated the absolute value of the time difference between the NCD and SLID peaks for each of the 536 waveforms in our dataset. The mean difference was 1.74 seconds. For 379 of the waveforms (70.7%), the difference between the two peaks was less than or equal to 1 second. SLID had a slight tendency to peak later than NCD, with later SLID peaks occurring for 291 (54.3%) of the waveforms. A plot of the differences is shown in Figure 25.

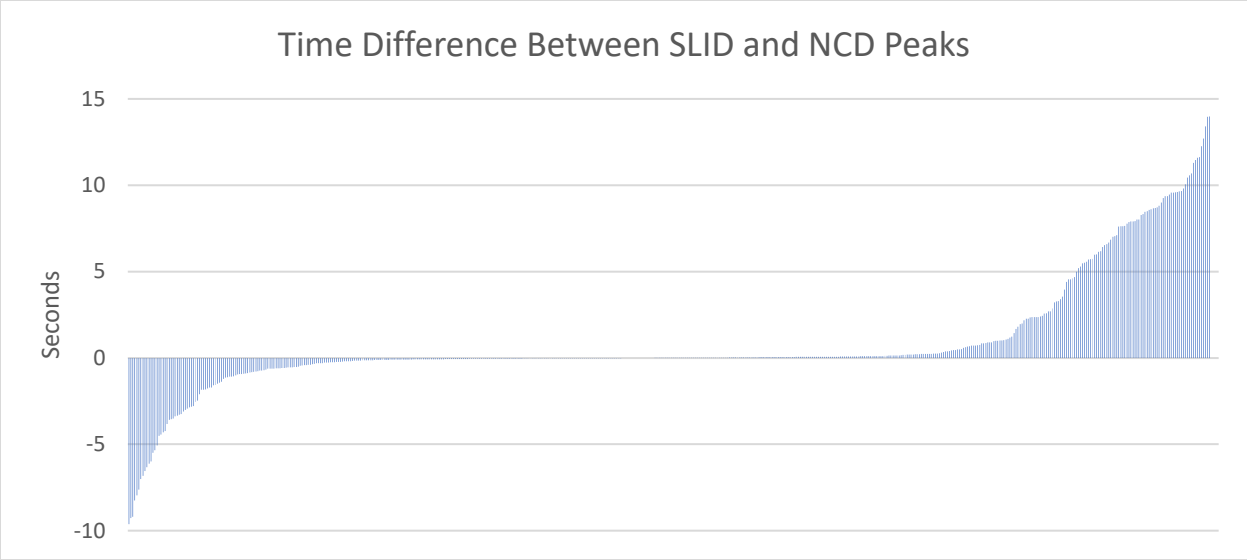

Figure 25. A plot showing the time difference between the maximum NCD peak and the maximum SLID peak for all 536 waveforms in our dataset.

The difference between the SLID peak and the expert analyst's best pick was slightly larger, on average, than it was for NCD. Across the entire set of waveforms, the average difference was 0.63 seconds for NCD and 1.65 seconds for SLID. This analysis took earlier peaks into account, when an earlier SLID or NCD peak was closer to the expert's best pick. Note that SLID and NCD differed in terms of when they produced earlier peaks that met our criteria (a peak at least 80% the height of the primary peak, occurring within 5 seconds before the primary peak). There were many waveforms for which NCD produced an early peak but not SLID, and vice versa. Figure 26 shows the difference between the SME's best pick and the closest NCD and SLID peaks. The waveforms are ordered by the size of the difference for SLID. This plot also highlights the tendency for SLID to peak slightly later than NCD. However, note that on the extreme ends of the plot, where the differences are large for SLID, NCD tends to have large differences as well, and often in the opposite direction. This indicates that those waveforms are very noisy and that neither algorithm is producing stable results. There are also many cases where SLID is much closer to the SME's pick than NCD, as illustrated by the bars in the center of the plot. Despite this, the results of the two algorithms are often quite consistent with one another.

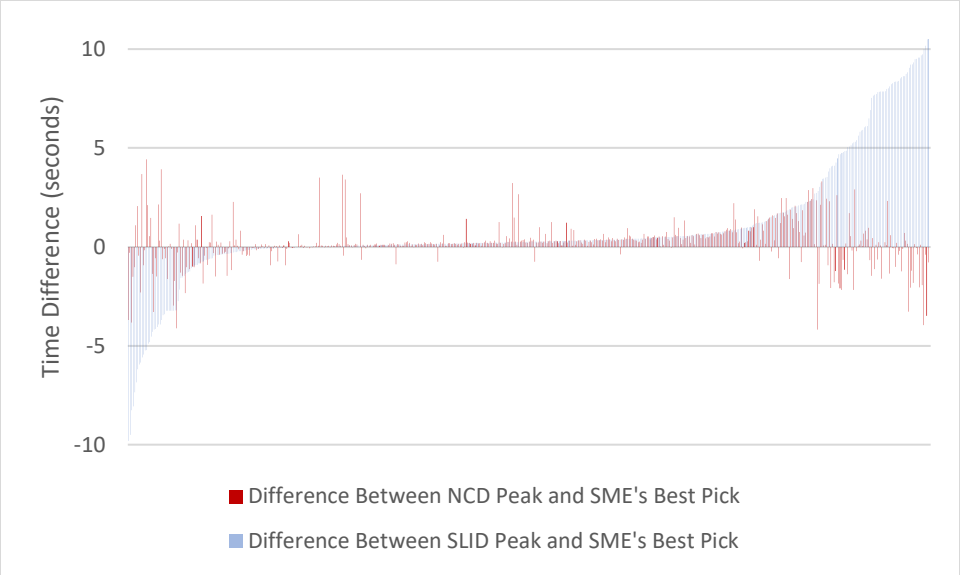

Figure 26. A plot of the differences between NCD and SLID and the SME's best pick for all 536 waveforms in the dataset.

A better way to compare the two algorithms is to select the best filter for each event. Recall that there are 100 events in this dataset, each of which was analyzed under multiple filter conditions to produce the 536 waveforms. As in our initial assessment of NCD, we used the difference scores to select the filter for which SLID performed the best for each seismic event. When selecting only the best filters, the performance of SLID and NCD was almost identical. The mean difference between the closest NCD peak and the expert's best pick was 0.47 seconds while the mean difference between the closest SLID peak and the expert's best pick was 0.54 seconds.

3.4.2. Comparing SLID to AIC and SME Picks

To compare SLID to AIC, we used a second dataset that contained SME picks for multiple seismic events and stations. The process for creating this dataset was similar to the process outlined for the dataset described in Section 2.3. The new dataset was intended to capture a wider variety of stations and events to provide a better picture of SLID's applicability to seismic analysis. While the first dataset consisted primarily of data collected from stations in Utah, the second dataset contained stations from all over the world. A map of the 37 stations included in the dataset is shown in Figure 27. There were a total of 98 waveforms selected for the dataset, each of which could be filtered using the 6 filters described in Section 2.3. The SME made picks for the earliest possible, most likely (best), and latest possible arrival times in every filter that had a useable signal. The result was a set of 444 waveforms.

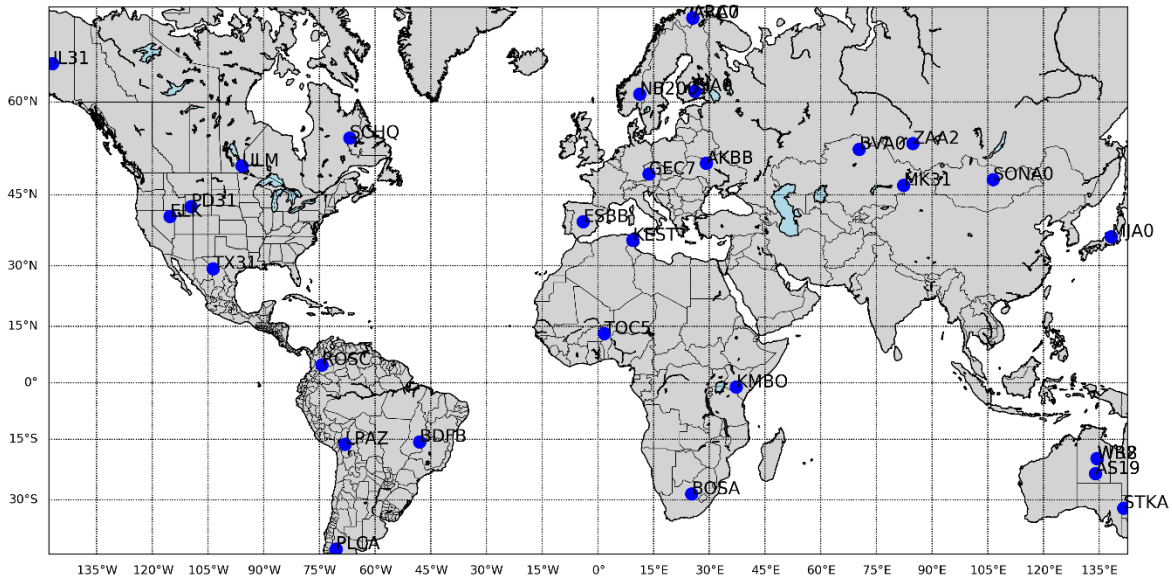

Figure 27. A map of the seismic stations included in the global dataset.

Across all 444 waveforms, the average difference between the maximum SLID value and the minimum AIC value was 6.99 s. The SLID peak was closer to the SME's pick than AIC for 240 of the 444 waveforms (54%). AIC tended to produce much larger errors (relative to the SME's pick) than SLID. On average, the SLID peak was 3.26 s away from the SME's best pick while AIC's pick was 6.68 s away from the SME's best pick. The results of the two algorithms were less than 1 second apart for 252 (56.8%) of the waveforms. When the waveforms where the results of the two algorithms were less than 1 second apart were excluded from the analysis, the SLID peaks were an average of 6.11 s away from the SME's best pick while the AIC picks were an average of 13.98 s away. SLID was closer to the SME's pick than AIC for 121 (63%) of the 192 waveforms in this group. These results show that the results of SLID tend to be more similar to the SME's picks than the results of AIC, particularly when the two algorithms produce results that are more than 1 second apart.

When looking at the best filter for each station/event combination (the filter for which SLID was closest to the SME's best pick), we found that the SLID peak was within 1 second of the SME's best pick for 87 of the 98 events. These results are shown in Figure 28.

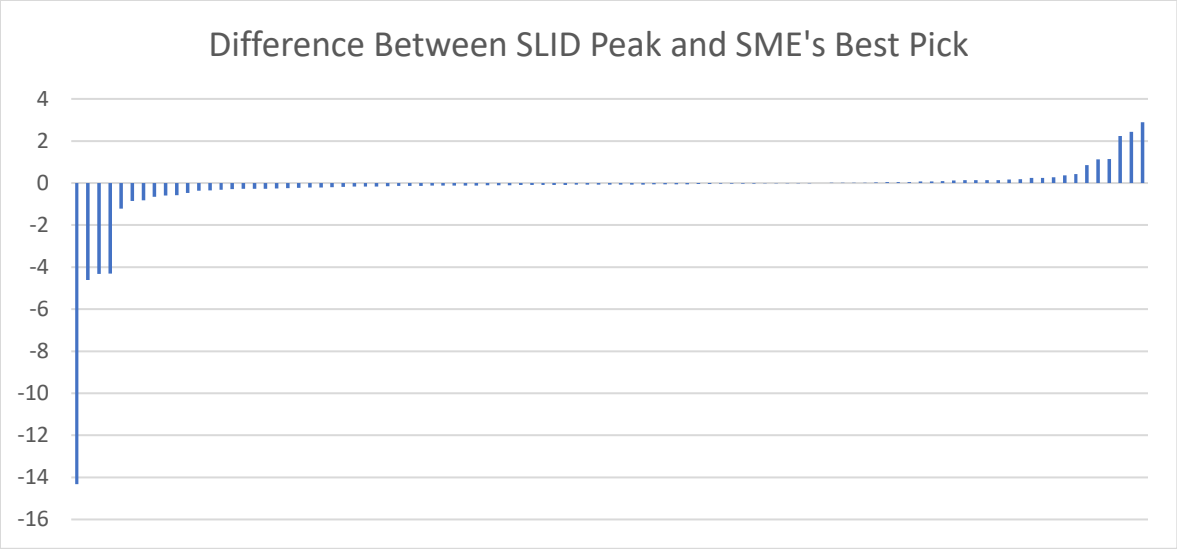

Figure 28. The difference between the SLID peak and the SME's best pick for the filter that produces best SLID result for each event/station combination.

This analysis indicates that the SLID peak is typically very close in time to the SME's pick for the most likely signal arrival time when an appropriate filter is chosen for the data. For the events where the time difference was large, such as the event at the far left side of Figure 28, the waveforms were very noisy and the SME could only discern a signal in one or two of the filters.

Given the computational advantages offered by SLID and the fact that it, like NCD, typically produces results that are very close to an SME's picks when the seismic waveforms are filtered appropriately, we chose to use SLID when developing code that implements this method for seismic analysts. SLID can be run in real time and offers the additional advantage of supporting a UQ analysis, as discussed in Section 3.3.

3.5. Assessing SLID for Minimizing False Detections

Another question when assessing the utility of compression metrics for seismic waveform analysis is whether these metrics can produce better automatic picks than existing methods. As discussed in the introduction, the existing automated processing method produces many extra picks. Human analysts end up discarding about 40% of the picks made by the system. Another way to reject these extraneous picks would be to determine whether or not they produce a SLID peak. To test this approach, we obtained a dataset consisting of 68 false detections that were generated in one day's worth of seismic data from one seismic station. For each of these false detections, the automated processing system placed a pick that was later rejected by a human analyst. We calculated SLID for a segment of data that included 60 seconds before and 60 seconds after each of the false detections. Most of the 68 examples did not have a clear SLID peak at all, and none of them had a peak that corresponded to the false detection. A representative example is shown in Figure 29. SLID was calculated using two different window widths for this example. Neither produces a peak near the time of the false detection, which is marked with a red line.

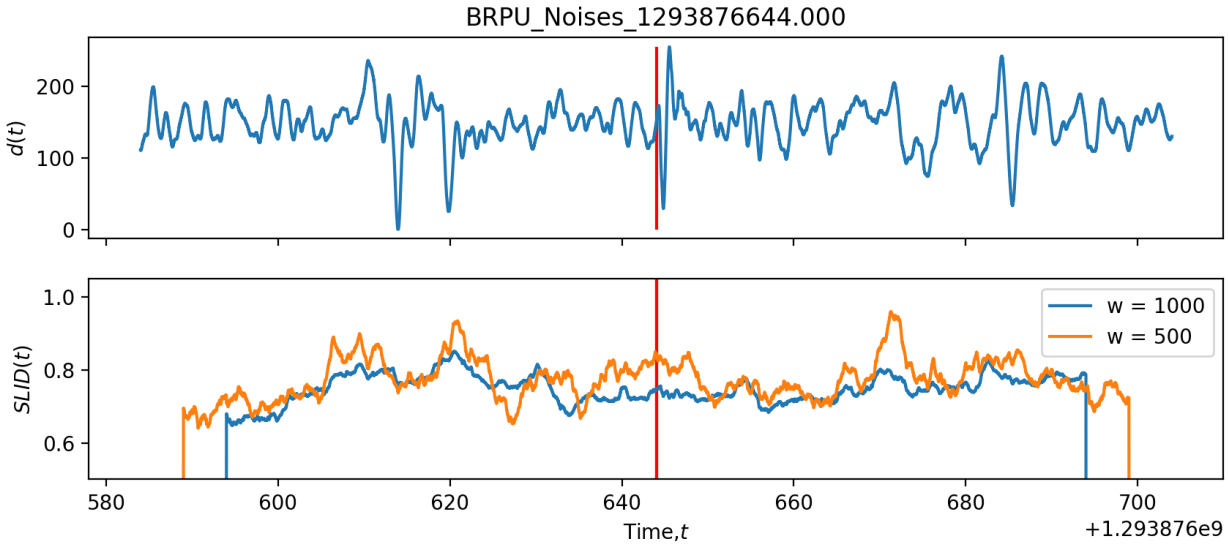

Figure 29. SLID plots for a representative example of a false detection made by the existing automated processing system.

This exploratory analysis indicates that compression metrics could be helpful for refining the event bulletins that are produced by the existing automated systems. The absence of a SLID peak could help analysts to reject the false detections more rapidly, or SLID could become a component of the automated processing system to provide converging evidence for or against the automatic picks.

This page left blank.

4. ASSESSING THE IMPACT OF INFORMATION FROM COMPRESSION METRICS ON HUMAN DECISION MAKING

4.1. Experimental Methods

4.1.1. Participants

This experiment was reviewed and approved by the Human Studies Board at Sandia National Laboratories. Eleven Sandia employees (6 male, 5 female) participated in this study and were compensated for their time. The average age of the participants was 45 and all of the participants had a master's or Ph.D. in geophysics or earth science (8 participants), electrical engineering (2 participants), or computer science (1 participant). All of the participants work in the seismic analysis and global nuclear monitoring domain and had extensive experience with seismic analysis.

4.1.2. Materials

4.1.2.1. Visualization Conditions

Nine different visualization conditions were developed for the experiment. One condition, as a baseline, presented only the seismic waveform. Another presented the waveform with a pick line based on the minimum AIC value for that waveform. This represents the current state-of-the-art, in which analysts see automatic picks based on the AIC algorithm. The other seven visualization conditions presented the SLID information in different ways. The SLID information could be presented as a plot of the scaled SLID values over time, a vertical line marking the maximum SLID value (visually equivalent to the pick lines based on AIC), or a highlighted range that indicated the location of the peak SLID values. The highlighted range started at 50% of the height of the peak on the left side of the peak and ended at 95% of the height of the peak on the right side of the peak. The highlighted range was biased toward the early/rising side of the peak due to our observation that SLID often peaks slightly late relative to SME picks. This highlighting system was intended to encompass the maximum SLID value as well as preceding values, indicating to the analyst where the arrival time is most likely to occur based on the SLID calculation. We presented these visual cues (SLID plot, SLID pick line, and highlighted range) individually and in every possible combination, as shown in Figure 30.

Based on our earlier research indicating that the maximum SLID peak is often slightly late relative to SME picks, we also marked earlier SLID peaks that met certain criteria. In order to be marked in the visualizations, these earlier peaks had to be a distinct local maximum that was at least 80% as tall as the tallest SLID peak, occurring within 5 seconds prior to the tallest SLID peak. The earlier peaks, when present, we marked with a lighter shade of blue than the main peak, as shown in Figure 30. The earlier peaks were also highlighted, with the highlighted range beginning at 95% of the height of the early peak on the left side of the peak and ending at the nearest local minimum on the right side of the peak. This highlighting is biased toward the late/falling side of the earlier peak, since the signal arrival times often occur close to or soon after the early SLID peaks. As with the highlighting for the dominant peak, the highlighting of the earlier peak is intended to draw the analyst's attention to the region of the waveform where the signal arrival is most likely to occur.

4.1.2.2. Stimuli

A total of 74 seismic waveforms were used as stimuli in the experiment. Two of these waveforms contained signals produced by known nuclear tests, one of which occurred in 1995 and one of

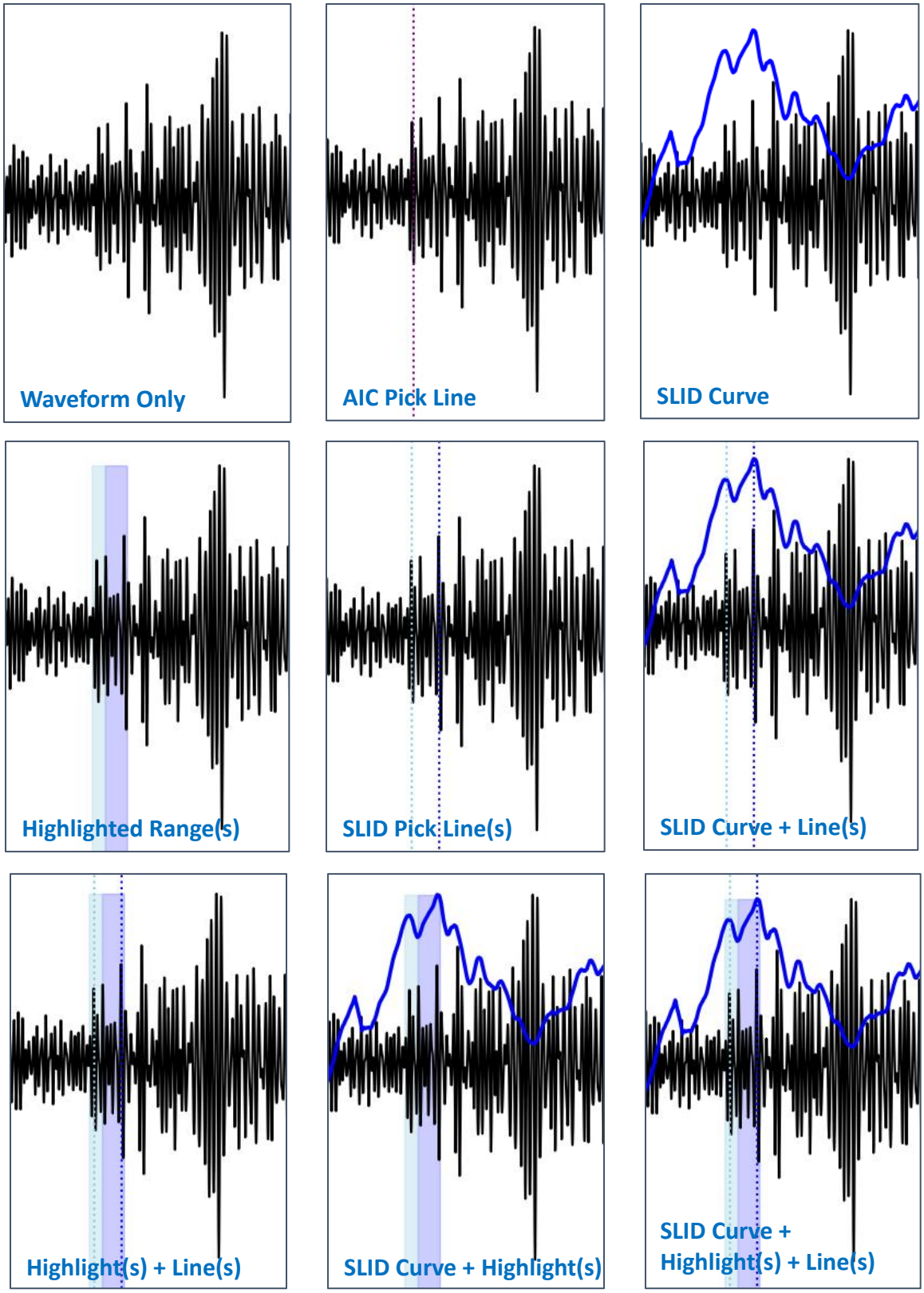

Figure 30. Examples of the nine visualization conditions used in the experiment.

which occurred in 2017. Those two waveforms were presented to all participants in all visualization conditions using a within-subjects design. The remaining 72 waveforms contained signals from a variety of seismic events, such as earthquakes and mining events. The waveforms were selected based on the following principles:

1) They were representative of common patterns that we had observed in the SLID outputs.
2) They varied in their visual representations (i.e., some had a wide highlighted range while others had a narrow highlighted range)
3) They emphasize cases in which the SLID results are not ideal (i.e., when there is not a single, sharp peak in the SLID plot). The current analytic workflow for the seismic analysts is highly effective, so it is important to demonstrate that adding new information to that workflow would not be distracting or detrimental to analyst performance, particularly in scenarios where the SLID peak is not sharp or does not coincide closely with the signal arrival time as determined by a SME. Noisy SLID plots occur when the waveform itself is noisy, yet those noisy waveforms represent the cases where adding information based on SLID is most likely to help (or hinder) the analyst's decision making. Since our goal is to use SLID to help analysts focus their attention on the most likely arrival times in the waveforms that are most difficult to analyze, it was important to include a high proportion of noisy waveforms that are difficult to analyze.

To meet these goals, we selected 8 waveforms that fell into each of the nine categories below. The "SME's best pick" refers to the time point that the SME selected as the most likely signal arrival time for that waveform. The "SME's range" refers to the time period defined by the SME's pick for the earliest possible and latest possible signal arrival times for each waveform. When that range was small, the SME was very certain about when the signal arrived at the seismic station. When that range was large, the SME was less certain. The nine categories of stimuli were as follows:

1) SLID had a sharp peak that overlapped tightly with the SME's range
2) SLID had a broad peak that overlapped tightly with the SME's range (i.e., both SLID and the SME's picks indicate the same band of uncertainty for the arrival times).
3) SLID had a broad peak that did *not* overlap with the SME's range
4) SLID peaked late relative to the SME's range
5) The SLID plot had an earlier peak (meeting our criteria for marking in the visualizations) and that earlier peak was a good match to the SME's best pick.
6) The SLID plot had an earlier peak (meeting our criteria for marking in the visualizations) but the larger peak was a good match to the SME's best pick.
7) The range of uncertainty indicated by highlighting the SLID peak(s) is much narrower than the SME's range.
8) The range of uncertainty indicated by highlighting the SLID peak(s) is much broader than the SME's range.
9) The SLID and AIC pick lines are very different from each other and from the SME's best pick.

All participants analyzed all 74 stimuli, but the pairing of the nine stimulus sets and the nine visualization conditions was counterbalanced across participants so that each participant saw each

stimulus set with only one of the visualization conditions. The stimulus set-visualization pairing and the order in which the participants completed the blocks was counterbalanced using a Latin Square design. The result was nine unique stimulus lists. Each list contained nine blocks of stimuli, one for each visualization condition. There were 10 waveforms in each block: 8 from one of the stimulus sets described above and the two stimuli that were presented in every visualization condition for all participants. The order of the waveforms in each block was randomized. Each of the waveforms was assigned a code (such as "A_01") so that the participants would be less likely to recognize any specific seismic events that they had seen or analyzed previously. In other words, the participants did not know when the signals depicted in each waveform occurred, which seismic station recorded the signal, or what kind of filtering had been applied to the waveform. They had to make their picks based only on the waveform itself and the information provided about SLID or AIC.

4.1.3. Procedure

Due to the COVID-19 pandemic, the experimental sessions were conducted via Skype. Participants digitally signed the consent form and returned it to the experimenter prior to beginning their session. After connecting with the experimenter via Skype, the participants were instructed on how to connect to a lab computer via a remote desktop connection. Once connected, the participants completed a demographic form and read a document that explained SLID and provided instructions for the task. Participants were then introduced to the graphical user interface (GUI) that was used to analyze the waveforms. They practiced using the GUI and making picks on two example waveforms.

When participants were ready to begin the experiment, they opened a checklist that indicated the order in which they were to analyze the waveforms and the visualization options that they should select in the GUI for each block. After opening a waveform, participants could zoom in on the waveform and pan back and forth as desired. They made picks for the earliest, best, and latest arrival times for the seismic signal. Each pick was mapped to a mouse button. One participant did not have access to a three-button mouse, so that participant only made picks to indicate the best arrival time. The participants could erase their picks and redo them as needed. When they were satisfied with their picks, they clicked the "save" button, closed the waveform, and checked it off on the list. The participant's picks were saved to a database, along with the time elapsed between when they opened the waveform and when they closed it.

The whole experiment, including training, took about 2-2.5 hours. Some participants broke it into multiple sessions, based on their schedules. Three of the participants completed the experiment in a single session, 7 broke it into two sessions, and one participant broke it into 3 sessions. There were a couple of cases where the data from a particular waveform did not save properly due to a bug in the code. If there were more than one or two trials with missing data, the participants were asked to log back in to re-analyze those specific waveforms.

4.2. Results

For each waveform, we calculated the width of the range of possible arrival times selected by the participant, as well as the time difference between the participant's best pick (their judgement of the most likely arrival time) and the SME's best pick, the SLID peak and highlighted range, the earlier peak and highlighted range (when applicable), and the AIC pick.

4.2.1. Participant Preferences

The participants were divided in terms of their preferences for the different visualization conditions. A subset of five participants felt that the conditions with the highlighted ranges were the most

helpful, while a different group of four participants thought that the highlighted ranges were the least helpful. The participants who liked the highlighting tended to dislike the vertical lines (both AIC and SLID lines) and the SLID plot. The participants who disliked the highlighting preferred the SLID plot or the SLID and AIC lines. Two participants said that the plot with everything on it (SLID plot, SLID lines, and highlighting) was their least favorite because it was too busy.

4.2.2. Within-Subjects Analysis

First, we analyzed the results for the two waveforms that showed seismic signals from nuclear weapons tests. Both waveforms were presented to every participant under all of the visualization conditions. The first of these waveforms, the event Lop Nor 1995 as recorded at seismic station SRU in Utah with a 0.8 high pass filter, was selected because it had a relatively clear signal. Figure 31 shows the waveform, zoomed in near the arrival time, with the best arrival time pick made by the SME, and the picks made by all 11 participants across the 9 visualization conditions (99 picks in total). Figure A-1 in Appendix A shows the full waveform, and Figure A-2 shows how the participants' picks were distributed in each visualization condition.

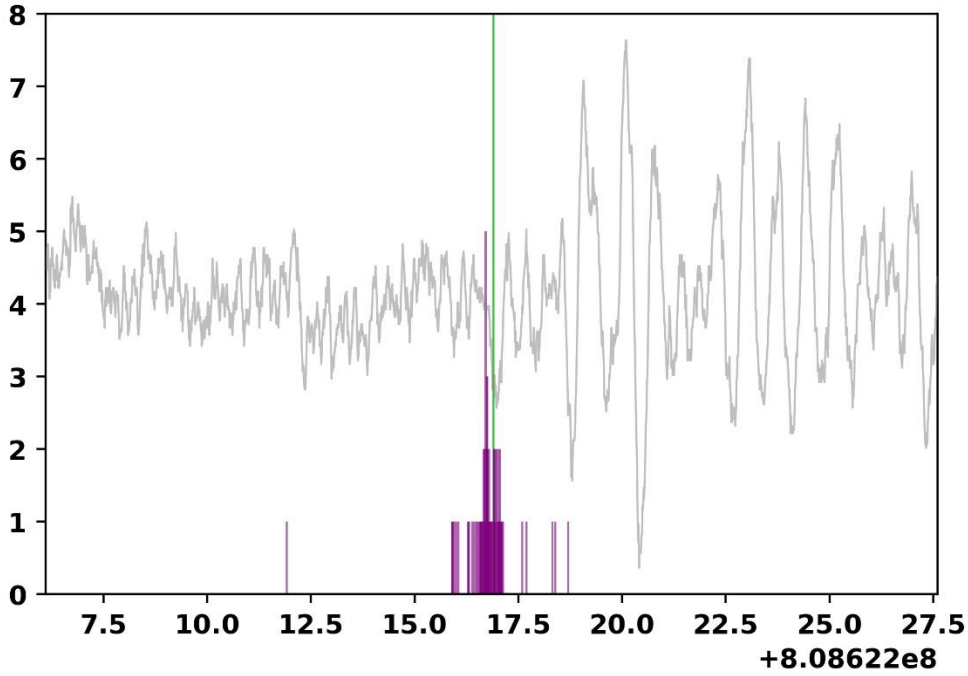

Figure 31. A zoomed-in view of the easier within-subjects seismic waveform (grey), the SME's pick for the best arrival time (green) and the participants' picks for the best arrival time (purple). The participants' picks are stacked if there were multiple picks falling in the same hundredth of a second. The y-axis indicates the number of picks in each time bin and the x-axis shows time in seconds.

Across all visualization conditions, there was only a 6.8 second range between the earliest time and the latest time selected by participants as the "best" arrival time. When we looked at the participants' choices for the best arrival time in each condition, the average values fell in a tight range, regardless of condition. These results are shown in Figure 32. Note that the entire time range depicted in the y-axis is only 2 seconds total. The average time of arrival selected by the participants fell within that same 2-second window, regardless of visualization condition.

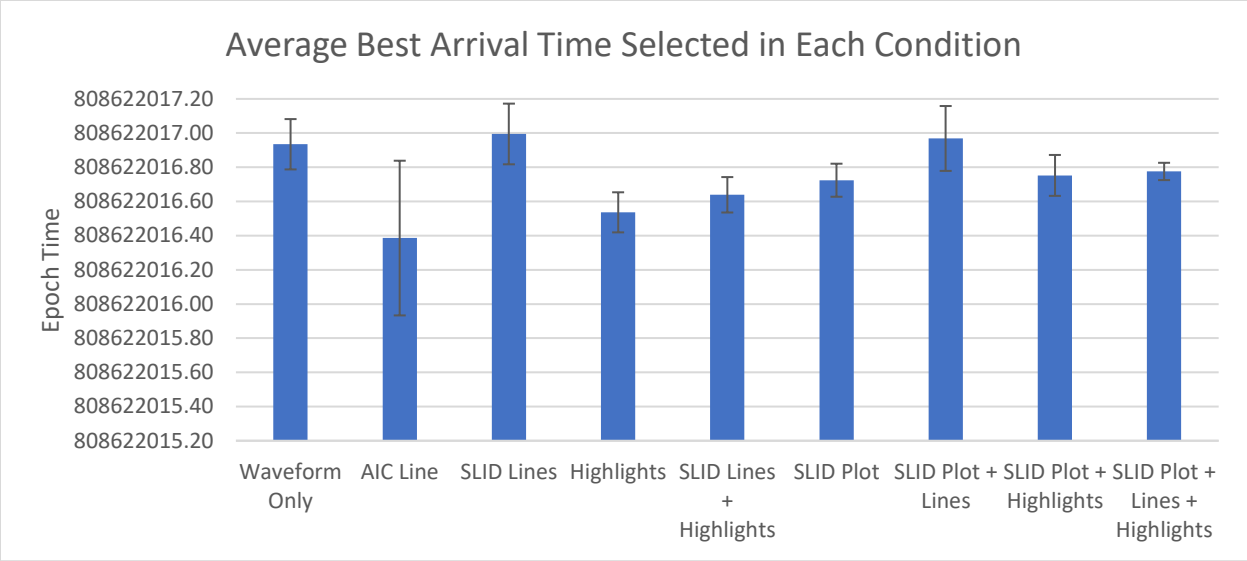

Figure 32. The average best arrival time selected for each visualization condition for the easier within-subjects stimulus.

Similarly, the participants' picks for the best arrival time were very similar to the SME picks that served as our point of comparison. The average difference for each visualization condition is shown in Figure 33. For all but one condition (the AIC visualization), the average difference was less than half a second.

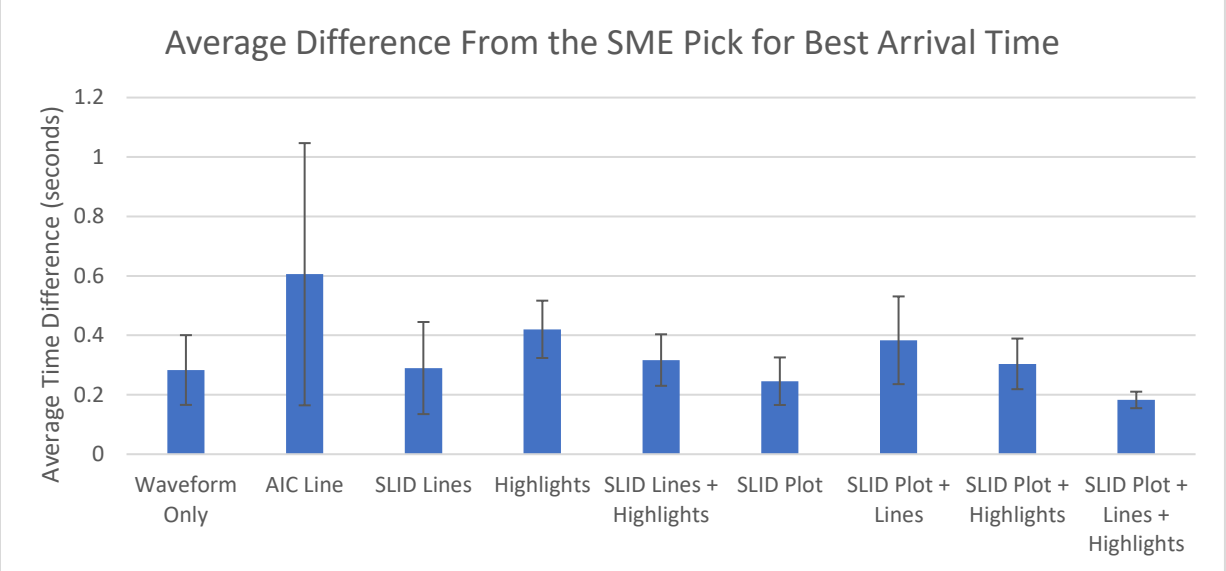

Figure 33. The average time difference in seconds between the participants' picks for the best arrival time and the SME's picks for the best arrival time in each visualization condition for the easier within-subjects waveform.

Unsurprisingly, there was not a statistically significant difference between the conditions. A repeated measures ANOVA showed that the visualization condition did not have a significant effect on the participants' accuracy ($F(8,80) = 0.48$).

The second nuclear test in the dataset was conducted in North Korea in September 2017 and the waveform was recorded at station SRU in Utah. A bandpass filter from 3.0-6.0 Hz was applied to the waveform used in our study. This waveform was selected because it was very noisy, particularly when participants first opened the file, prior to zooming in on time points of interest. We predicted that there would be more variability in the participants' picks across visualization conditions for this noisier, more difficult waveform. Figure 34 shows the waveform, the SME's best pick, and the picks made by the study participants across all visualization conditions. The full waveform is shown in Figure A-1 in Appendix A. Figure A-3 in Appendix A shows how the participants' picks were distributed in each visualization condition.

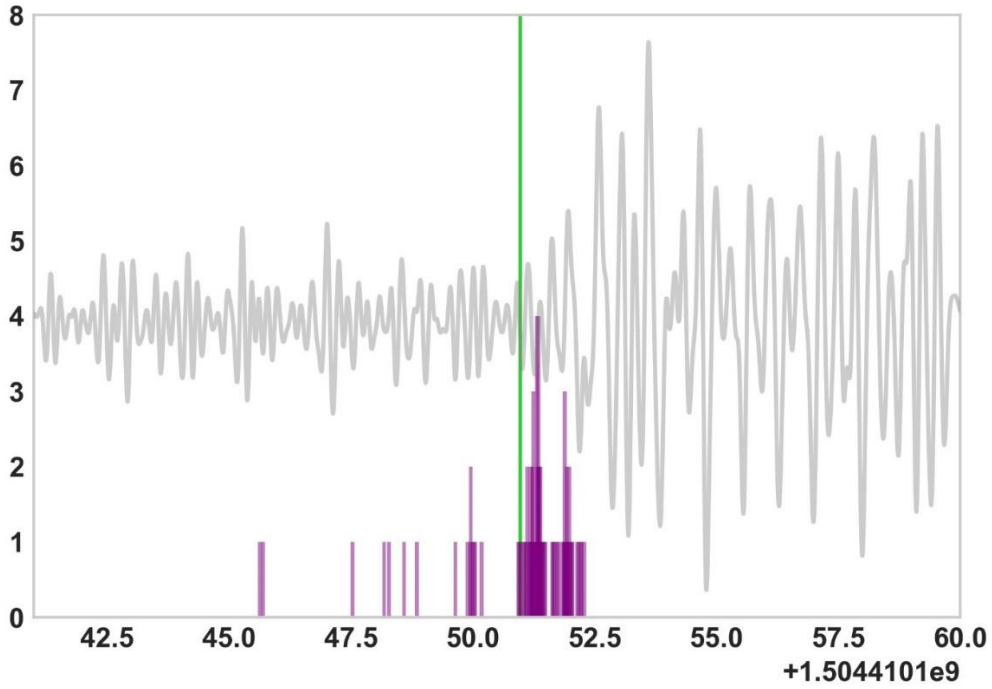

Figure 34. A zoomed-in view of the more difficult within-subjects seismic waveform (grey), the SME's pick for the best arrival time (green) and the participants' picks for the best arrival time (purple). The participants' picks are stacked if there were multiple picks falling in the same hundredth of a second. The y-axis indicates the number of picks in each time bin and the x-axis shows time in seconds.

There was a much larger range of pick times for this waveform than for the easier stimulus discussed above. In this case, the participants' best picks spanned a range of 95.2 seconds. However, this large span was mostly driven by one participant who picked a time that was much earlier than the other participants' picks. Note that this early pick is not shown in Figure 34. Without this outlier, the remaining picks spanned a range of 6.7 seconds, which was very similar to the range observed for the easier waveform. The average time for the participants' best picks in each visualization condition are shown in Figure 35 (the very early outlier is excluded from this analysis). Once again, note that the time range on the y-axis spans only 3 seconds. On average, the picks were quite close together.

However, it is notable that the variability in the picks is much larger for the Waveform Only and SLID Lines conditions than in the other conditions.

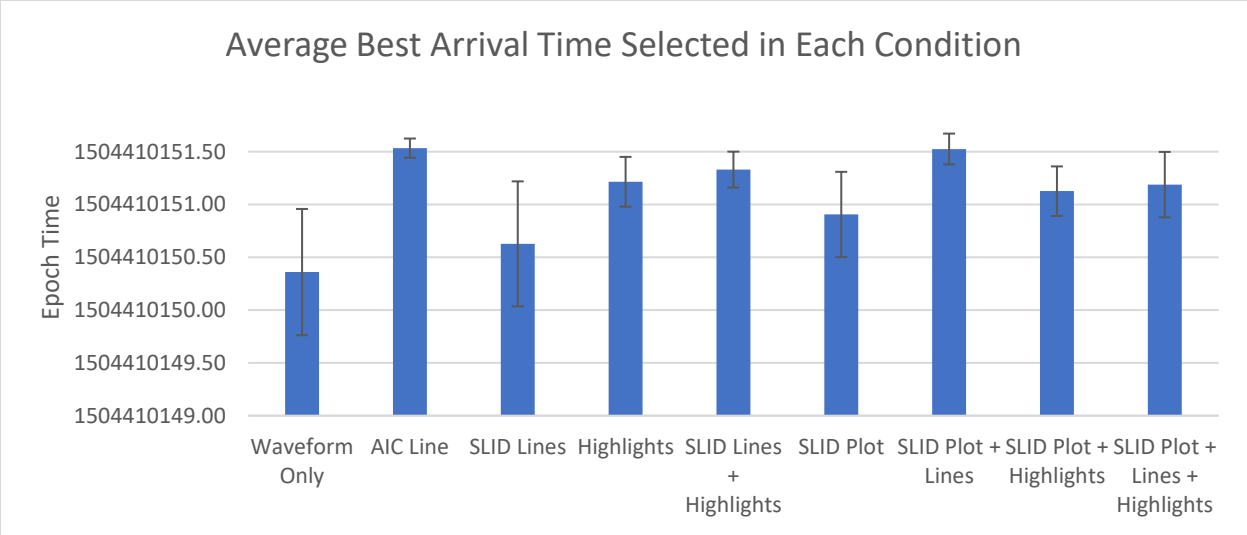

Figure 35. The average best arrival time selected for each visualization condition for the more difficult within-subjects stimulus.

The average difference between the participants' picks for the best arrival time and the SME's picks for each visualization condition is shown in Figure 36. The Waveform Only condition had the largest average difference and the highest variability. However, a repeated-measures ANOVA showed that there was not a significant effect of visualization condition on the participants' accuracy relative to the SME ($F(8,79) = 1.06$, $p = 0.4$).

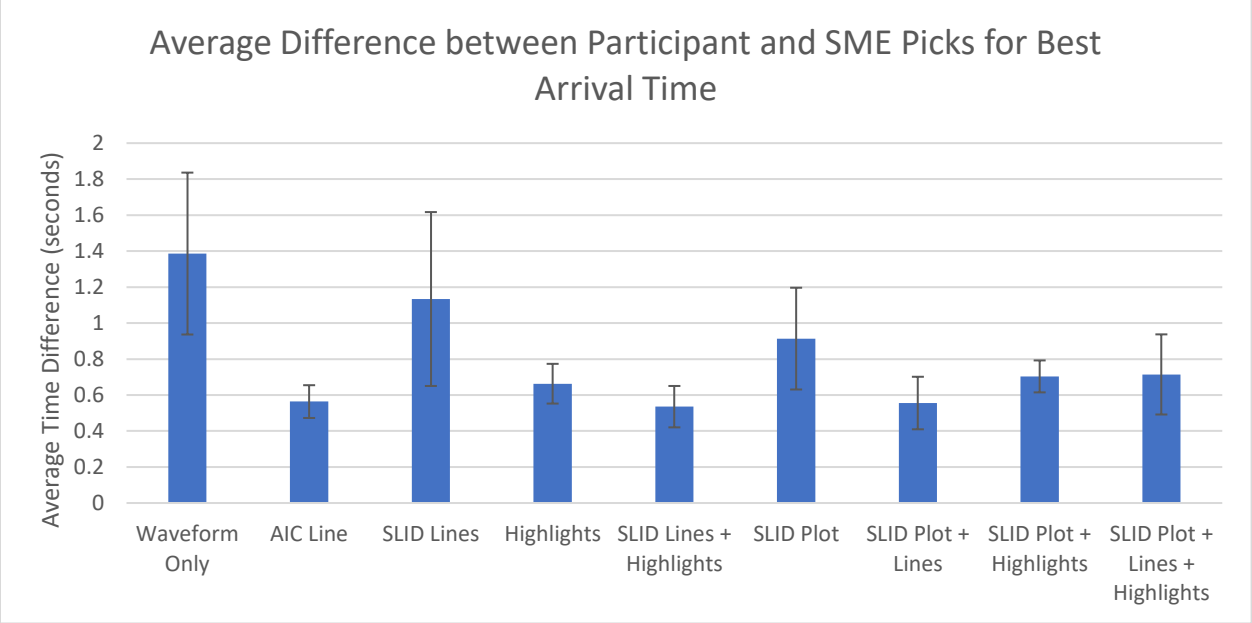

Figure 36. The average time difference in seconds between the participants' picks for the best arrival time and the SME's picks for the best arrival time in each visualization condition for the more difficult within-subjects waveform.

4.2.3. Between Subjects Analyses

For most of the waveforms in the experiment, participants analyzed the waveforms in only one visualization condition. For these waveforms, we used a between-subjects analysis to assess the impact of the visualization conditions on the participants' responses. Our primary goal was to test whether the information based on SLID would be a hinderance to the participants when it was not a good match to the SME's pick. In other words, in those situations, did participants make picks that were closer to the SME's pick or closer to the times indicated by SLID?

As a first step in assessing the impact of SLID information on the participants' picks, we calculated the difference between each participant's best arrival time pick, the best arrival time picked by the SME, and the time of the maximum SLID value. The results for all waveforms are shown in Table 1. The bolded cells indicate the instances where the outputs of each algorithm were displayed.

Table 1. Mean (and Standard Deviation) Difference Results for All Between-Subjects Stimuli

Condition	Average Difference Between Participant's Best Pick and SME's Best Pick	Average Difference Between Participant's Best Pick and the Time of the Peak SLID Value	Average Difference Between Participant's Best Pick and the Time of the AIC Pick
Waveform Only	1.10 (SD = 2.10)	3.45 (SD = 20.23)	3.20 (SD = 20.26)
AIC Line	0.86 (SD = 1.06)	2.83 (SD = 20.18)	**2.65 (SD =20.19)**
SLID Line(s)	1.02 (SD = 1.13)	**3.10 (SD = 20.08)**	2.88 (SD = 20.09)
SLID Highlight(s)	0.90 (SD = 1.01)	**4.88 (SD = 28.41)**	4.77 (SD = 28.43)
SLID Line(s) + Highlight(s)	0.94 (SD = 1.28)	**4.78 (SD = 28.29)**	4.80 (SD = 28.31)
SLID Plot	1.00 (SD = 1.21)	**3.06 (SD = 20.13)**	3.08 (SD = 20.12)
SLID Plot + SLID Line(s)	2.20 (SD = 9.23)	**3.86 (SD = 22.24)**	3.96 (SD = 22.28)
SLID Plot + Highlight(s)	0.90 (SD = 1.07)	**2.99 (SD = 20.29)**	3.02 (SD = 20.30)
SLID Plot + SLID Line(s) + Highlight(s)	1.00 (SD = 1.91)	**2.92 (SD = 20.08)**	2.82 (SD = 20.08)

These results show that the participants' best picks were closer to the SME's best picks than they were to the SLID or AIC picks, regardless of visualization condition. This is a good indication that providing the information based on SLID did have a detrimental impact on the participants' performance.

To investigate this further, we analyzed the stimuli where the difference between the SME pick and the maximum SLID peak was more than 1 second. This was the case for 21 of the 72 waveforms that were used in the between-subjects stimulus set. For those waveforms, the average difference

between the participants' picks and the SME's picks was 1.37 s (SD = 1.42 s) when only the waveform was shown and 1.67 s (SD = 1.93 s) when the SLID information was shown. The difference between the participants' picks and the time of the SLID peak was 9.22 s (SD = 37.10 s) when only the waveform was shown and 9.78 s (SD = 40.34) when the SLID information was shown. In other words, displaying information based on SLID did not lead the participants to make picks that were closer to the SLID peak than they were to the signal arrival time, even when the SLID peak was relatively inaccurate. This is the situation in which adding additional information is most likely to mislead users, but we did not observe that happening in this study.

When the difference between the SLID peak time and the SME's pick was less than 1 second, participants benefited from the SLID information, making picks that were closer to the SME's pick. For this subset of waveforms, the average time between the participants' picks and the SME's picks were 0.99 s (SD = 2.32 s) when only the waveform was shown. This difference dropped to 0.90 s (SD = 4.21 s) when the SLID information was shown. The participants' picks also moved closer to the SLID peak in these situations. The average difference between the participants' picks and the SLID peak was 1.08 s (SD = 2.32 s) when only the waveform was shown, but dropped to 0.93 s (SD = 4.23 s) when the SLID information was shown. While this is a small effect, it indicates that participants recognize when SLID aligns with their interpretation of the waveform and place their picks closer to the SLID peak. In contrast, when the SLID information did not align well with the SME's picks, the participants successfully ignored the SLID peak and placed their picks close to the SME's pick, regardless of where the SLID peak occurred.

It is worth noting that several of the participants commented that they do not trust automatic picks in seismic waveforms, regardless of the algorithm doing the picking. One participant commented that "the best algorithms used in seismology are only accurate 50% of the time, so I don't trust it." Since all of our participants had extensive experience with seismic waveform analysis, they may be predisposed to ignore information supplied by algorithms when it does not match with their own assessments. It is possible that displaying SLID results that do not match well with expert judgements would be more problematic for less experienced analysts.

It is also important to note that this experiment did not include any UQ information for SLID. We would expect the UQ information to improve accuracy for experienced and novice analysts alike, by allowing them to assess how confident they should be in the SLID information. If the UQ analysis shows with high confidence that a pick should be in a particular spot, we would expect to see the same effects that we observed in this study, where participants made picks that were close to the SLID peak when it provided them with a "good" result. In contrast, if the UQ analysis indicates low confidence, that provides analysts with information that would help them to ignore the SLID information and rely on their own knowledge instead. If less experienced analysts are more likely to be negatively impacted by unstable SLID outputs, we would expect the UQ output to counteract any negative impact. It is also likely that providing analysts with UQ information would help them to calibrate their trust in the algorithm's outputs appropriately.

4.3. Summary of User Study Results

The results of this study indicate that participants used the SLID information when it aligned well with the SME's judgements, and that this information helped them to make more accurate picks. They also recognized the cases when the SLID information was not particularly helpful (cases when the SLID peak was more than 1 second away from the SME's best pick). In those cases, the participants successfully ignored the SLID peak and made picks that were much closer to the SME's pick than they were to the SLID peak.

We did not observe any significant effects of visualization condition in the within-subjects analysis. Instead, we found that about half of the participants had a strong preference for seeing highlighted ranges to help direct their attention to a particular section of the data. The other half of the participants preferred to see the SLID plot and vertical pick lines that are similar to the way AIC picks are typically displayed.

The findings of this study support the way in which SLID and UQ information are displayed in the SLIDPick Graphical User Interface, which is described in Section 5. When using the SLIDPick tool, users can choose whether or not to display the SLID plot and the UQ information. Like the highlighted ranges, the UQ information provides the users with cues about the time periods in which the signal arrival is most likely to occur. It is superior to the highlighting used in the user study, which was based on somewhat arbitrary cutoffs related to the peaks on a single SLID calculation. The UQ information incorporates numerous iterations of SLID, so it can both guide the user's attention to the appropriate regions and indicate how much confidence they should have in any given SLID plot.

This page left blank.

5. GRAPHICAL USER INTERFACE FOR APPLYING SLID TO SEISMIC WAVEFORMS

One of the key outcomes of this project is the development of the Sliding Information Distance Pick (SLIDPick) graphical user interface (GUI) that allows users to compute SLID with UQ for any seismic waveform. This section provides a brief overview of the features of the GUI.

5.1. Computing Environment

The SLIDpick GUI is written in the Python programming language, version 3.6 and will run on Windows, Mac, and Linux operating systems - provided the supporting Python modules are installed on the given platform. The program has been successfully tested in a Python 2.7 environment with the necessary packages present. A list of required Python packages is shown in Table .

Table 2: SLIDpick Required Python Modules

Required Python Packages		
pandas	subprocess	os
time	tkinter	matplotlib
string	sqlite3	pylab
sys	obspy	scipy
datetime	math	itertools
numpy	getpass	basemap

5.1.1. User-Specific Parameters

The SLIDpick GUI requires the user to define three parameters that are dependent upon the user's installation. A SLIDpick.par file defines the location of these parameters and must be located in the same folder as the SLIDpick software. For directory paths, ensure that the trailing slash is included in the variable name.

Table 3: SLIDpick Parameters

Parameter	Description
SLID_db	sqlite3 database file location
Waves_loc	Directory of seismic waveforms, if analyzing data that is stored locally
projection_lib	Map information required for event and bulletin maps

5.2. Using the Interface

5.2.1. Initiating the Program

Based on operating system, there are different methods to initiate the program.

- On Linux, include the environment on the first line of the SLIDPick_v1.1.py script (e.g., *#!/usr/bin/env /path/to/Python /env)*
- For Windows, first activate the Python environment and then call the program.

source activate environment name
Python SLIDpick_v1.1.py

After issuing one of the commands above, the SLIDpick GUI is launched (Figure 37).

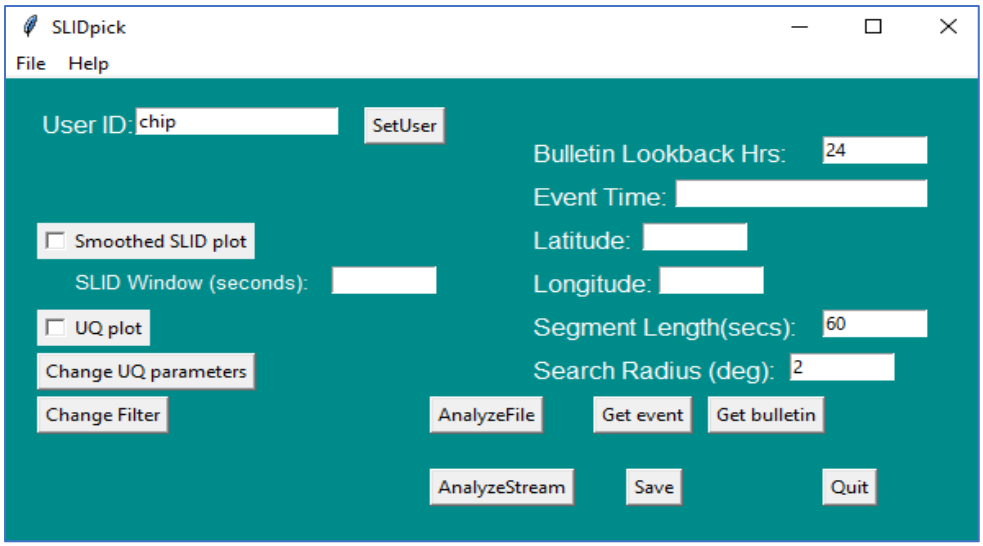

Figure 37. SLIDpick User Interface.

- File->Open: This option will invoke a file selection dialog that lists waveforms stored in the Waves_loc directory discussed in Section 5.1.1. The "Analyze File" button invokes the same function.
- File->Save: This option, invoked after processing, will write the user's results to the SLID_db. The "Save" button invokes the same function.
- File->Exit: This option allows the user to exit the program and ends the user's session. The "Quit" button invokes the same function.
- Help dropdown menu: Opens a window with usage instructions.
- SetUser button: Enter a unique User ID (see Section 5.2.2 for further instructions for using this field).
- Smoothed SLID plot button: Compute and display SLID attributes in a waveform analysis window.
- SLID window entry form: Allows user to set number of seconds used for SLID window computation.
- UQ plot button: Allows the user to apply and display UQ refinement processing
- Change UQ parameters button: Allows user to change parameters for UQ processing.
- Change Filter button: Invokes a filter selection dialog where the user changes the data filter. There are six available filters, which are the standard filters used in seismic analysis, as discussed in Section 2.3.
- Bulletin Lookback entry form: Allows the user to set time range, in hours, from current time to retrieve an event bulletin from IRIS.
- Event Time entry form: Allow the user to input an event origin time to be used for real-time event-based waveform retrieval from IRIS. A date and time are required, *e.g.* 2020-09-15T00:00:00, where time is in Universal Coordinate Time (UTC).

- Latitude entry form: Allows user to input an event latitude for event based real-time waveform retrieval.
- Longitude entry form: Allows user to input an event longitude for event based real-time waveform retrieval.
- Segment Length entry form: length of waveform segment past the origin time to be retrieved.
- Search Radius entry form: radius from the given latitude and longitude of an event to search for stations and retrieve waveform data.
- AnalyzeFile button: Invokes a file selection dialog that lists waveforms from the acquired experiment datasets stored in the *Waves_loc*.
- Get Event button: Retrieves waveforms from the IRIS data center using the origin time, latitude, longitude, segment length and search radius as input parameters. Returns all available BHZ, HHZ, EHZ, and SHZ components from stations within the search radius.
- Get Bulletin button: retrieves the event bulletin from IRIS for events occurring between the present time minus the Bulletin Lookback hours. Results in a popup dialog that allows the user to select an event, show on a map, and retrieve waveforms.
- Analyze Stream button: After the user retrieves an event via the Bulletin popup, this button will sequentially loop through the newly acquired waveforms, apply SLID and/or UQ plots, and allow the user to analyze the data stream.
- Save button: This option, invoked after processing, will write the user's results to the *SLID_db*.
- Quit button: Allows the user to exit the program and end the user's session.

5.2.2. Setting the User ID

Prior to beginning a session, the user should set an identifying label in the "User ID" text entry widget. The User ID will be given to the analyst by the person conducting the experiment. The ID 000 is set by default and processing will not be allowed until a suitable ID is entered and the "Set User" button depressed (Figure 38). The ID can be set to any alpha-numeric string, based on the preference of the experiment administrator. After the User ID is set, processing can begin.

Figure 38. User ID field used for setting the unique alpha-numeric ID using the SetUser button.

5.2.3. Reading Waveforms into the Interface

After the User ID is set, the user can click on the "AnalyzeFile" button to open a waveform file for processing. A dialog of files (in the *Wave_loc* directory) is opened.

After selecting a waveform file from the dialog shown in Figure 39, the waveform is displayed in the analysis window (Figure 40.)

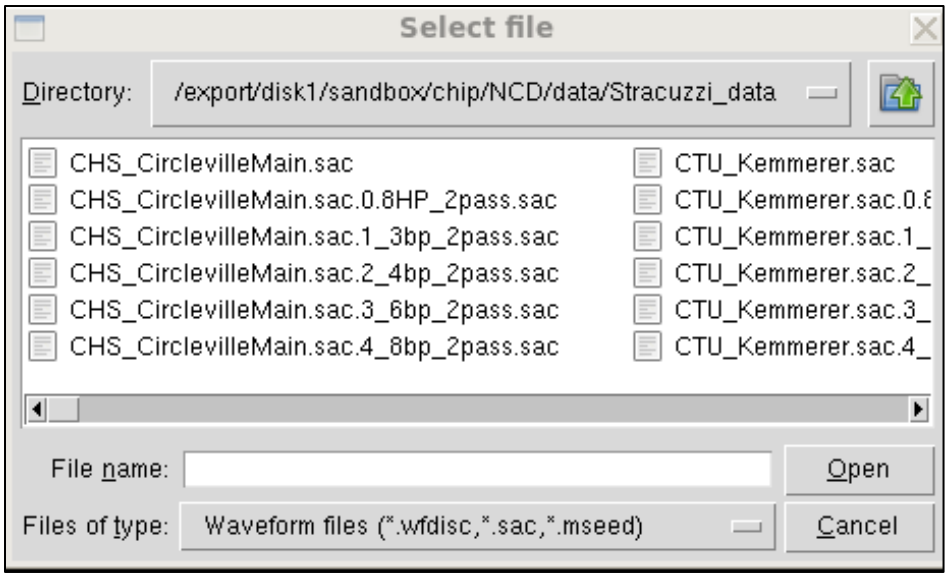

Figure 39. Waveform file selection dialog.

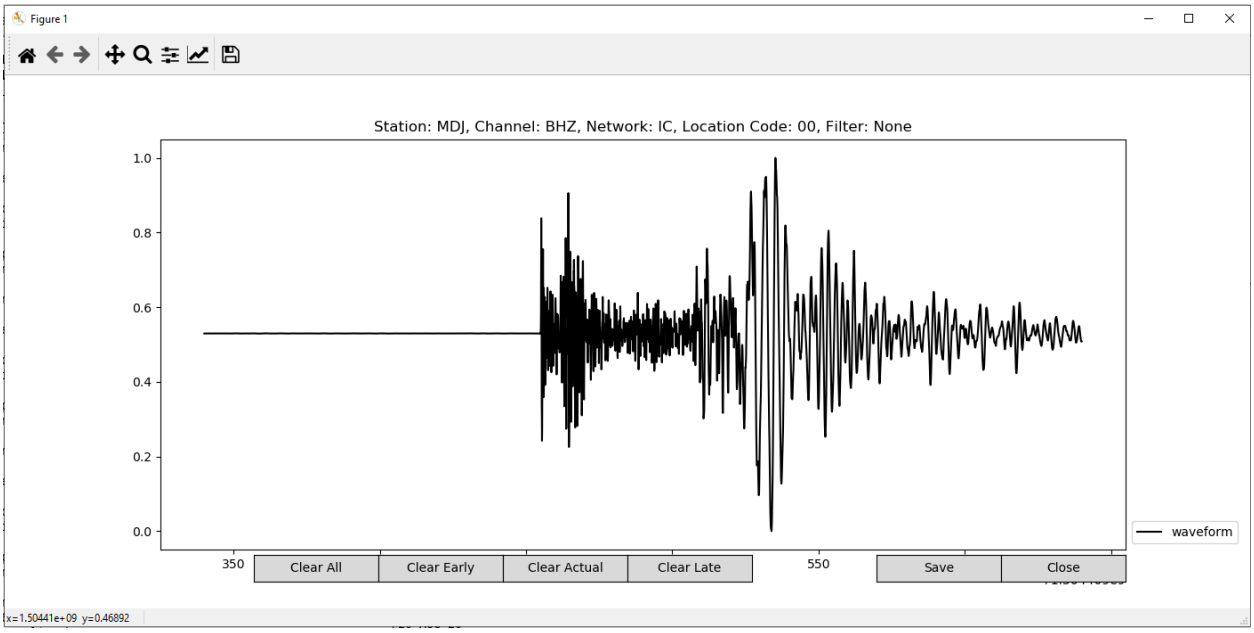

Figure 40. The waveform analysis window.

5.2.4. Displaying Features on Waveforms

Figure 41 shows buttons and entry forms on the left side of the SLIDpick GUI. The buttons are labeled Smoothed SLID plot, SLID Window, UQ plot, Change UQ parameters, and Change Filter. Any combination of these can be selected to display attributes and set parameters for the waveform analysis window. The attributes to be displayed are computed at processing time.

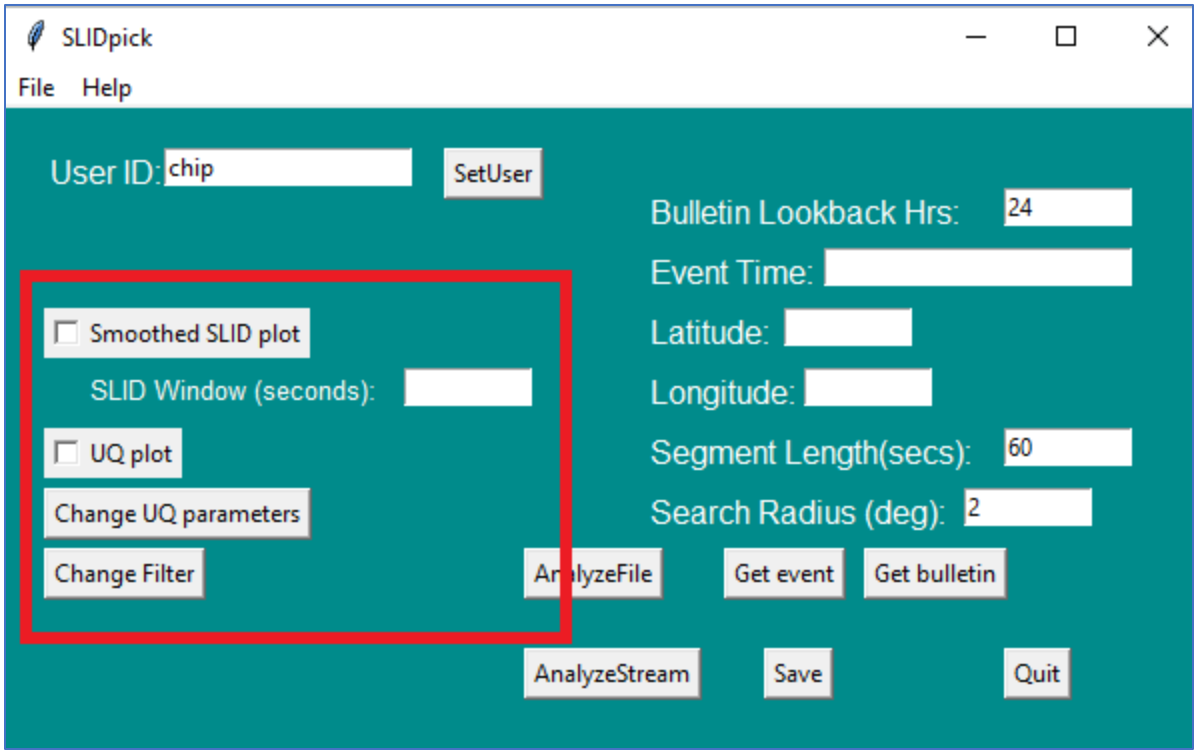

Figure 41. Location of Smoothed SLID plot, UQ plot, UQ parameters, and Change Filter buttons on the SLIDpick screen.

- The Smoothed SLID plot button will calculate SLID for each waveform using the default parameters discussed earlier in this report. The user can change the default window size by using the SLID Window entry box.
- The UQ plot button will show the distribution of the maximum SLID scores over a range of parameters.
- The Change Filter button allows the user to select one of five high pass or bandpass filters, as described above.

The Uncertainty Quantification can be used to assess the stability of the SLID result. The combination of attributes displayed will be logged in the database when analysis results are saved. Figure 42 illustrates a waveform with all attributes displayed.

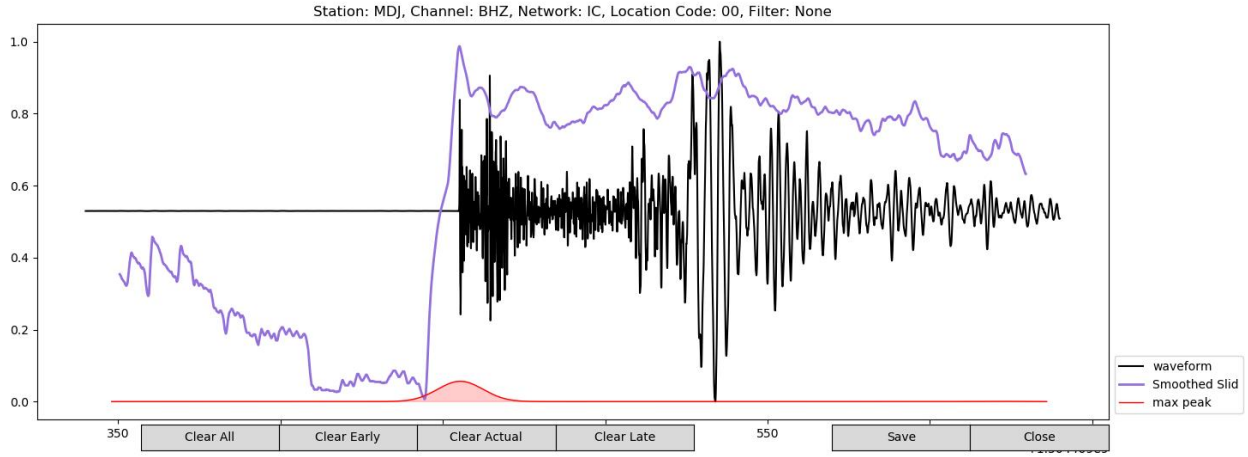

Figure 42. Waveform with SLID and UQ features plotted. The SLID plot using the default parameters is shown in purple and the UQ result is depicted by the red distribution plot along the bottom edge of the window.

5.3. Using Get Bulletin Interface to Retrieve Data

If a user wants to analyze seismic data from the Incorporated Research Institutions for Seismology (IRIS) database, rather than waveforms saved locally on their computer, they will use the Get Event and Get Bulletin features. The Get Bulletin button in the GUI activates the Bulletin interface Figure 43. This allows the user to retrieve a bulletin from IRIS and select events from the bulletin for processing.

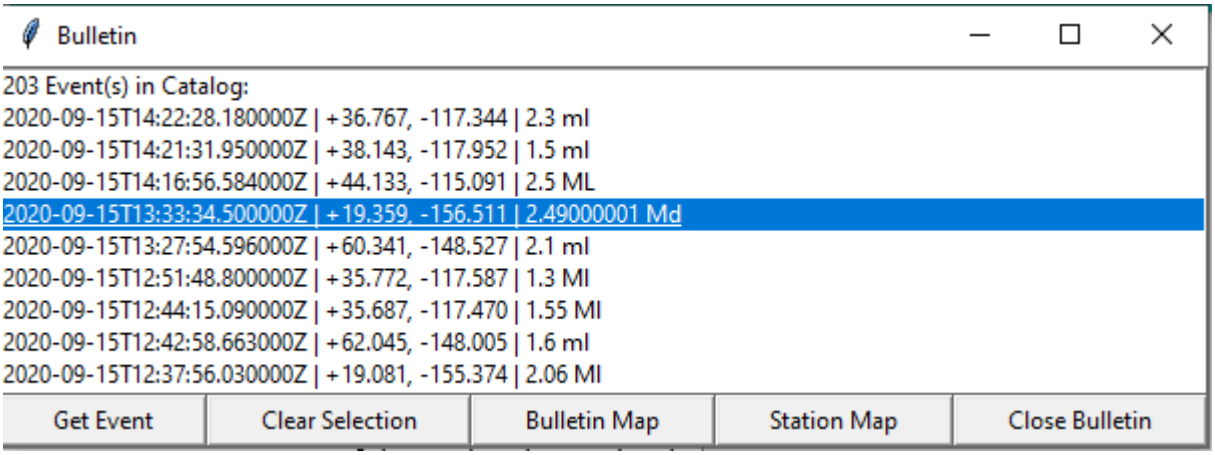

Figure 43. The Bulletin interface.

The bulletin interface shows the list of events pulled from IRIS for the time period specified by the user in the main GUI. Users can select an event from the list and use the Get Event button to retrieve the data for analysis. If users want to explore the events available in the list, the Bulletin Map button will produce a global map of events in the bulletin, as shown in Figure 44.

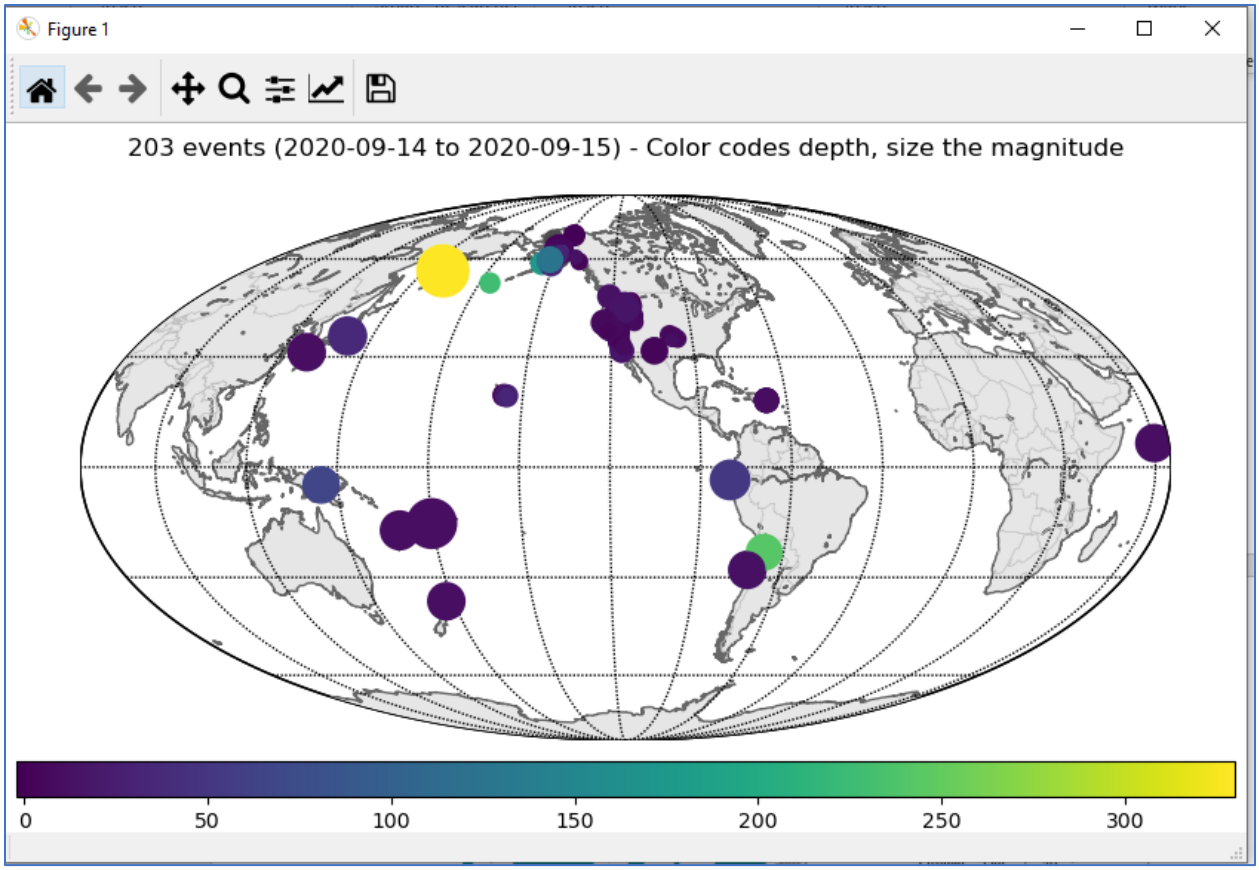

Figure 44. Bulletin Map

The Station Map button will plot stations within the radius of the event. This button can be particularly useful prior to using the Get Event button because it can inform the user how many stations will be requested. An example is shown in Figure 45. Users can then change the search radius if they want to retrieve data from more stations (or fewer stations). As the search radius increases, the segment length should be increased as well, to account for the fact that the signal will arrive later at more distant stations. Clicking on the Get Event button will retrieve available waveforms from stations within the search radius entered on the main window of the event. The waveforms will start at the origin time and end at the origin time plus the segment length time. The Clear Selection button will clear the current selection. The Close Bulletin button simply closes the dialog.

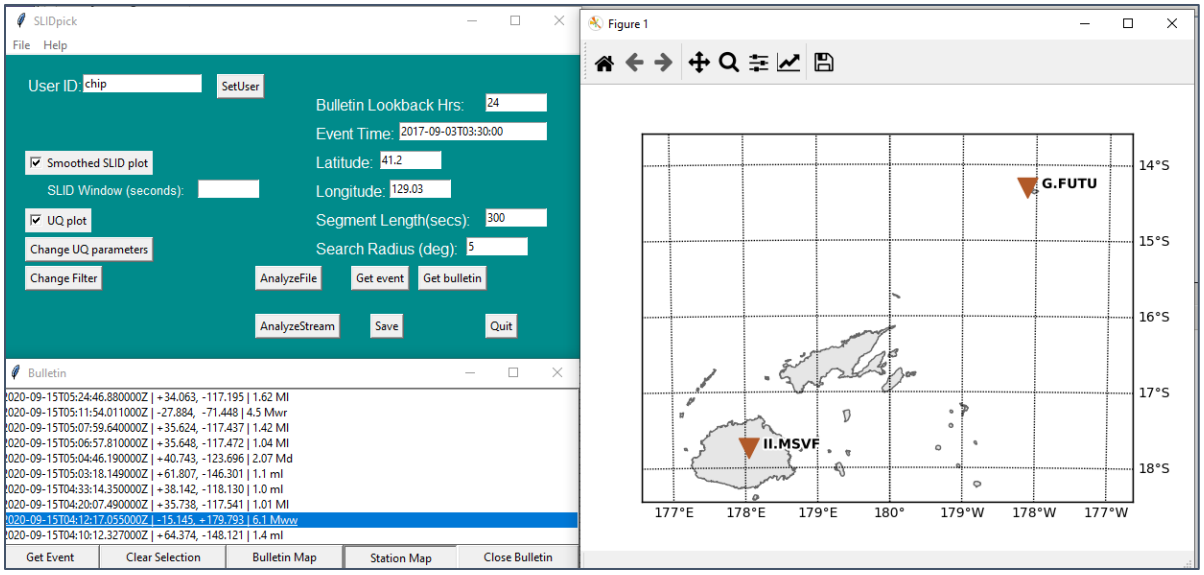

Figure 45. Bulletin dialog with event selected and Station Map displayed for stations within 5 degrees of the event. Labels contain the network and station code.

With the event selected in Figure 46 the result of pressing the Get Event button is shown in Figure 47.

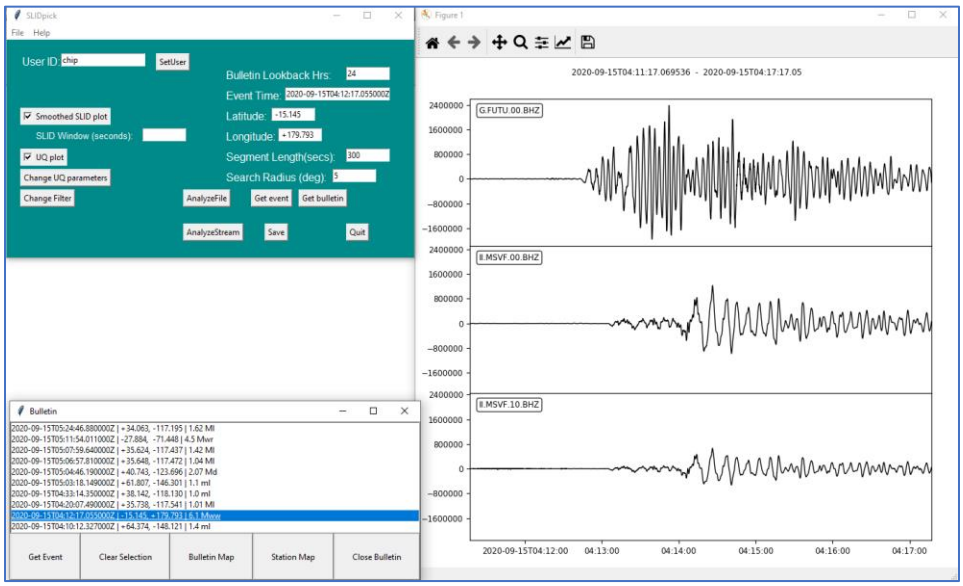

Figure 46. Waveforms retrieved with Get Event button.

After the user has retrieved the data using the Get Event button, clicking on the Analyze Stream button in the main window will process the data trace-by-trace and bring each up in the analysis window, as shown in Figure 47.

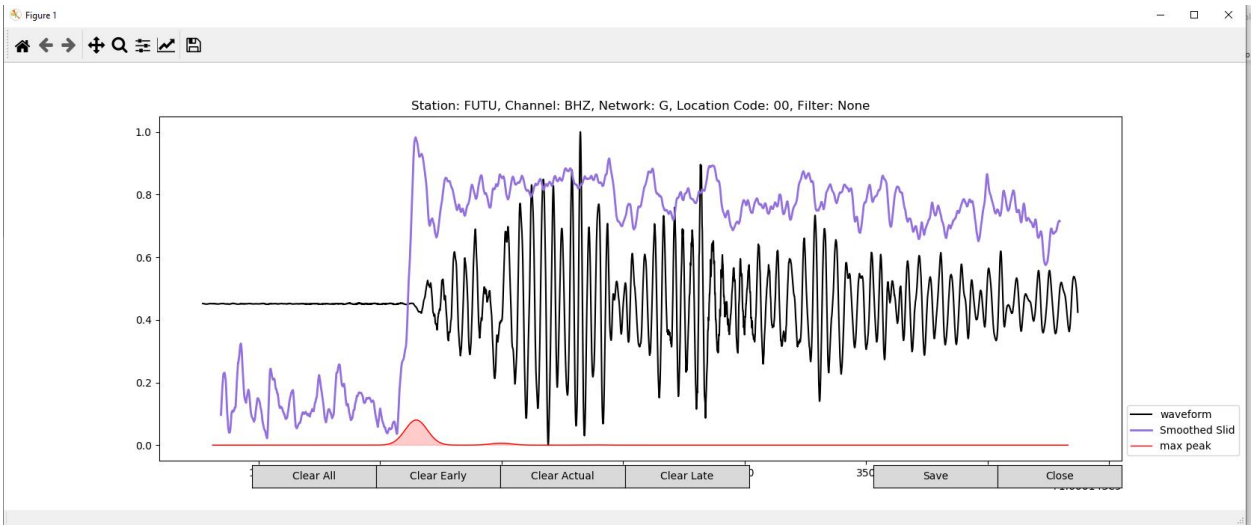

Figure 47. Data for station FUJU, for an event at 12:17:05 UTC on 09/15/2020, near the Fiji Islands located at latitude: 15.145 S and longitude: 179.793 E.

5.4. Analysis

Analysis, or start time picking, occurs in the waveform analysis window. After the user has set their ID, selected attributes to display, and opened a waveform, they are ready to make their earliest, actual, and latest seismic phase arrival time onset estimates.

In the window, the user will notice an orange vertical line appear when the mouse cursor is in the waveform widget. The orange line is for reference. To pick the earliest seismic phase arrival time onset, slide the orange line to the appropriate place on the waveform and click the left mouse button, a blue dashed vertical line will appear.

To mark what is believed to be the actual seismic phase arrival time onset, click the middle mouse button (or scroll wheel depending upon the mouse) and the time will be marked by a vertical green line.

The right mouse button will mark the latest seismic phase arrival time onset and the resulting dash-dot vertical line will be shown in wheat.

If the user has a change of opinion and would like to change one of the estimates, select the corresponding "Clear" button on the bottom left side of the analysis window (see Figure 48).

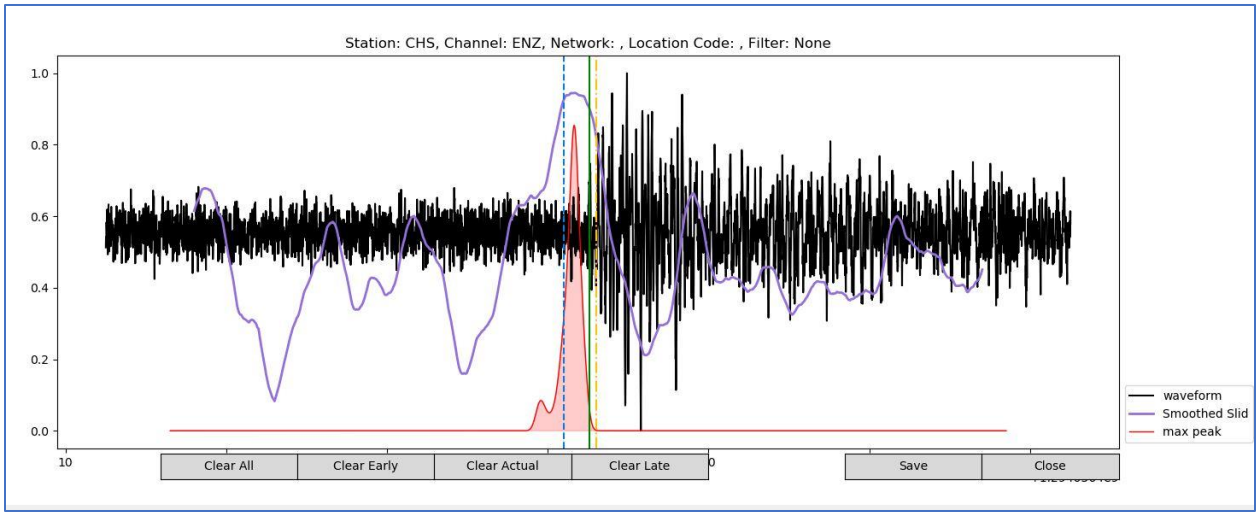

Figure 18. Analysis window with analyst onset estimates.

In Figure 48, all three picks are shown on the waveform as dashed blue, solid green, and dash-dot wheat vertical lines. Once the picking has been accomplished, the user can click on either the "Save" or "Close" buttons on the bottom right portion of the analysis window. Clicking "Save" will write the results to the database, this button calls the same function as the "Save" button on the main form of the GUI or the File->Save menu option. If the results are saved from this interface, the GUI will issue an error message dialog if the user tries saving from using either of the other options. The "Close" button will shut the waveform window. If needed, results can be saved using one of the other methods after the waveform window is closed.

If a closer look is needed the user can 'zoom' in on the waveform using the magnifying glass icon in the tool bar of the analysis window (Figure 47). Note that while the zoom is active, the cursor for analysis is disabled. A second click of the magnifying glass is required to re-activate the cursor.

The units on the y-axis in the upper plot in Figure 18 are normalized amplitude. Time is plotted on the x-axis in units of seconds.

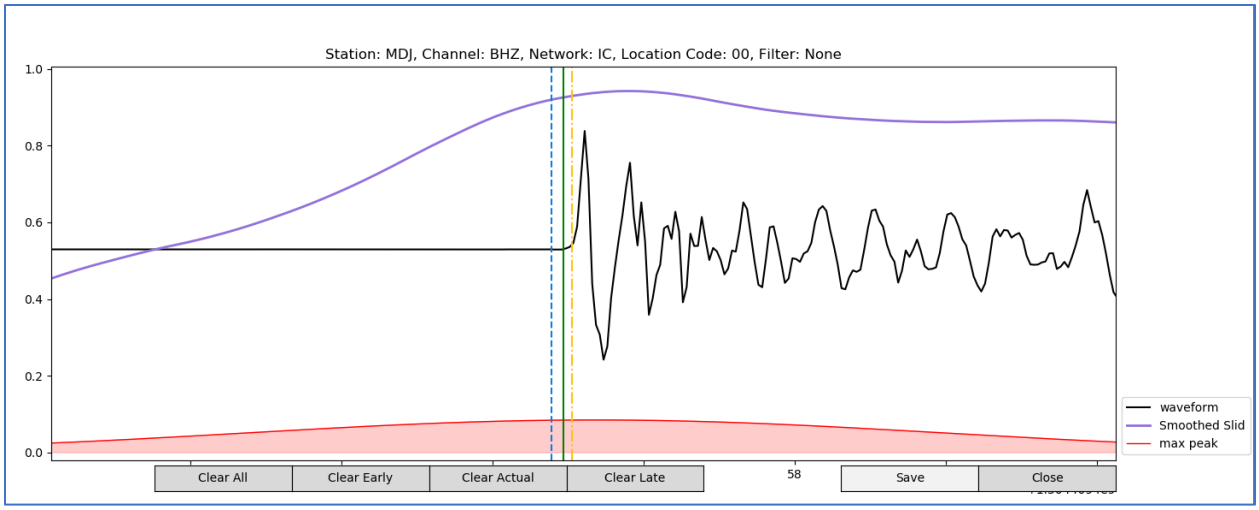

Figure 49. Zoomed view of annotated waveform shown in Figure 48.

5.5. Database

The SLID database is an SQLite Version 3 database that consists of two tables. To query the database type: *sqlite3 SLIDpicks_db*

The database will open, and the user will have a *sqlite>* prompt. *Type Entering .schema* at the prompt will reveal the table creation syntax and show the attributes of each table.

The SLID_picks table (Table 1) contains the name of the *waveform* and *user*, epoch and human readable times for each onset estimate, a *proc_time_seconds* value that indicates how long the user took to make their picks, a *condition* value that reflects which attributes (if any) were displayed while the user was processing the waveform, and a *lddate* field showing the date/time the entry was made.

Table 1: The SLID_picks Database Table

Field	Attributes	Description
Waveform	text	Waveform Name
User	text	User Identification
Earliest_time_epoch	real/float	Epoch time of earliest start
Actual_time_epoch	real/float	Epoch time of actual start
Late_time_epoch	real/float	Epoch time of latest start
earliest	text	Human readable earliest start time
actual	text	Human readable actual start time
latest	text	Human readable latest start time
Proc_time_seconds	integer	Analyst processing time
condition	text	Overlays/attributes applied during analysis
Lddate	text / date	Date/Time record saved to the database

The slid_ref table (Table 2) is a mapping between the waveforms in the dataset and the associated csv files containing the AIC and SLID attributes that are plotted if selected. Typical SQL queries can be made of each table, for example, typing:

select user,waveform,earliest,latest from slid_picks
where waveform='CHS_CirclevilleMain.sac.1_3bp_2pass.sac'

will yield the result:

002|CHS_CirclevilleMain.sac.1_3bp_2pass.sac|2011-01-03 12:07:21.610011|2011-01-03 12:07:22.568235

Table 2: The SLID_ref Database Table

Field	Attributes	Description
Slid_no	text	Cross-reference for numbered SLIDcsv files
waveform	text	Waveform name

5.6. Obtaining the Code

The code for running the GUI is hosted at https://gitlab.sandia.gov/SeismicNCD For access, please contact the first author of this report.

6. CONCLUSIONS

This project demonstrated that NCD and SLID can be applied to seismic waveforms to support global nuclear explosion monitoring. Although both methods are effective for detecting signal arrival times, we focused on SLID because it can be computed several orders of magnitude faster than NCD. This fast computation time allows for UQ analyses that indicate how confident the user should be in any picks based on SLID. Our user study demonstrated that providing information based on SLID was helpful to analysts, particularly for noisy waveforms. A key concern was whether SLID would distract analysts and lead to less accurate analyses when the SLID result did not correspond well to a SME's analysis of a waveform. To address this concern, we emphasized cases in which a single SLID calculation produced mediocre results (note that poor SLID results obtained from a single calculation would be mitigated by using the UQ analysis as well). Even in this worst-case scenario, we did not find any evidence that SLID misled the analysts or produced cognitive biases.

SLID could be used to support seismic analysis in a variety of ways, such as the following:

1) Corroborating the picks made by AIC. If the two algorithms agree, it is very likely that the pick is a true positive and that the timing of the pick is accurate. Conversely, when there is no SLID peak corresponding to an automatic pick, that pick could be a false positive and warrants further inspection by an analyst.
2) Selecting the optimal filter band for displaying a waveform. Effective filters will enhance the visibility of the seismic signal and result in a SLID plot with a sharp peak and high confidence.
3) Helping analysts to direct their efforts more efficiently. If a pick is well-supported by the information content of the waveform, as indicated by SLID, there is no need to spend time adjusting that pick. However, when SLID produces a broad peak, a peak that rises gradually, or a UQ analysis that indicates medium or low confidence, these are all indications that expert judgment is needed for making an accurate pick. The analysts bring external knowledge to bear on the problem that allows them to make accurate picks even for noisy signals that are not conducive to automated analyses. SLID can help to identify those signals so that analysts are applying their expertise when it is needed and not wasting time on cases where the automatic picks are sufficient. In addition, even when the SLID result has low or medium confidence according to the UQ analysis, both the SLID plots and the UQ plots can help to direct the analyst's attention to the most likely time regions for the signal arrival.

Future directions for this research could include testing the impact of the UQ analyses on analyst performance as well as investigating ways to incorporate the SLID and UQ calculations into the analysts' workflow to support the applications described above.

7. REFERENCES

[1] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," Knowledge and Information Systems, vol. 51, pp. 339–367, 2017.

[2] A. Bardera, M. Feixas, I. Boada, and M. Sbert, "Compression-based image registration," in 2006 IEEE International Symposium on Information Theory. IEEE, 2006, pp. 436–440.

[3] D. Cerra and M. Datcu, "A fast compression-based similarity measure with applications to content-based image retrieval," Journal of Visual Communication and Image Representation, vol. 23, no. 2, pp. 293–302, 2012.

[4] H. Koga, Y. Nakajima, and T. Toda, "Effective construction of compression-based feature space," in 2016 International Symposium on Information Theory and Its Applications (ISITA). IEEE, 2016, pp. 116–120.

[5] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitányi, "The similarity metric," IEEE Transactions on Information Theory, vol. 50, no. 12, pp. 3250–3264, 2004.

[6] M. Li and P. M. B. Vitányi, An Introduction to Kolmogorov Complexity and its Applications, 2nd ed. Springer-Verlag, 1997

[7] A. Macedonas, D. Besiris, G. Economou, and S. Fotopoulos, "Dictionary based color image retrieval," Journal of Visual Communication and Image Representation, vol. 19, no. 7, pp. 464–470, 2008.

[8] McNamara, L. A., Bauer, T. L., Haass, M., & Matzen, L. (2016, October). Information Theoretic Measures for Visual Analytics: The Silver Ticket? In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (pp. 53-61).

[9] E. Raff and C. Nicholas, "An alternative to NCD for large sequences, Lempel-Ziv Jaccard distance," in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.

[10] E. Stamatatos, "A survey of modern authorship attribution methods," Journal of the American Society for information Science and Technology, vol. 60, no. 3, pp. 538–556, 2009.

[11] C. L. Ting, R. V. Field, Jr., T.-T. Quach, and T. L. Bauer, "Generalized boundary detection using compression-based analytics," in IEEE International Conference on Acoustics, Speech, and Signal Processing, May 12–17 2019, pp. 3522–3526

[12] Ting, C. L., Fisher, A. N., & Bauer, T. L. (2017, September). Compression-based algorithms for deception detection. In *International Conference on Social Informatics* (pp. 257-276). Springer, Cham.

[13] P. M. B. Vitányi, F. J. Balbach, R. L. Cilibrasi, and M. Li, "Normalized information distance," in Information Theory and Statistical Learning. Springer, 2009, pp. 45–82.

[14] S. Wehner, "Analyzing worms and network traffic using compression," Journal of Computer Security, vol. 15, no. 3, pp. 303–320, 2007.

[15] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," IEEE Transactions on information theory, vol. 23, no. 3, pp. 337–343, 1977.

[16] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," IEEE transactions on Information Theory, vol. 24, no. 5, pp. 530–536, 1978

APPENDIX A. SUPPLEMENTAL MATERIALS

Table A-1. The Combination of Events and Stations Used in the NCD Evaluation Dataset

Event Name	Stations Selected
Bald Mountain	DUG, FLU, FSU, PSUT
Bingham 2	DBD, DCM, IMU, MLI, MMU, NMU, OWUT, PTU, SRU BHZ, SRU EHZ, WCU
Bingham Canyon (earthquake)	LHUT, MLI, MMU, NAIU, PNSU, RDMU, WCU
Chile (earthquake)	SRU BHZ, SRU EHZ, SRU HHZ
Circleville AS (earthquake)	FSU, HLJ, LCMT, LHUT, PNSU
Circleville Main (earthquake)	CHS, FSU, HLJ, HONU, LCMT, LHUT, PNSU, PTU
Kemmerer	CTU, MLI, MSU, NAIU, NLU, PTU, RBU, RCJ, SRU, WVUT
Longwall Large (mining)	DCM, PNSU, ROA, SRU BHZ, SRU EHZ, SRU HHZ
Longwall Small (mining)	PNSU, ROA
Lop Nor 1995 (nuclear test)	INCN, SRU EHZ
NK Sep 2017 (nuclear test)	INCN, MDJ, SRU EHZ
Peabody	FLU, MMU, MSU, SNO, SRU BHZ, SRU EHZ, SRU HHZ, WCU
Salt Lake	CWU, DAU, DUG, FLU, FSU, LEVU, PTU, WBC
Sufco (mining)	ELU, EMU, FSU, HTU, LEVU, SNO, SRU, TCRU, TMU ENZ, TMU HHZ, WCU
Tohoku (earthquake)	DCM, DUG, INCN, PD31, SRU BHZ, SRU EHZ, SRU HHZ
Vanuatu (earthquake)	FSU, ICU, SRU BHZ, SRU EHZ, SRU HHZ

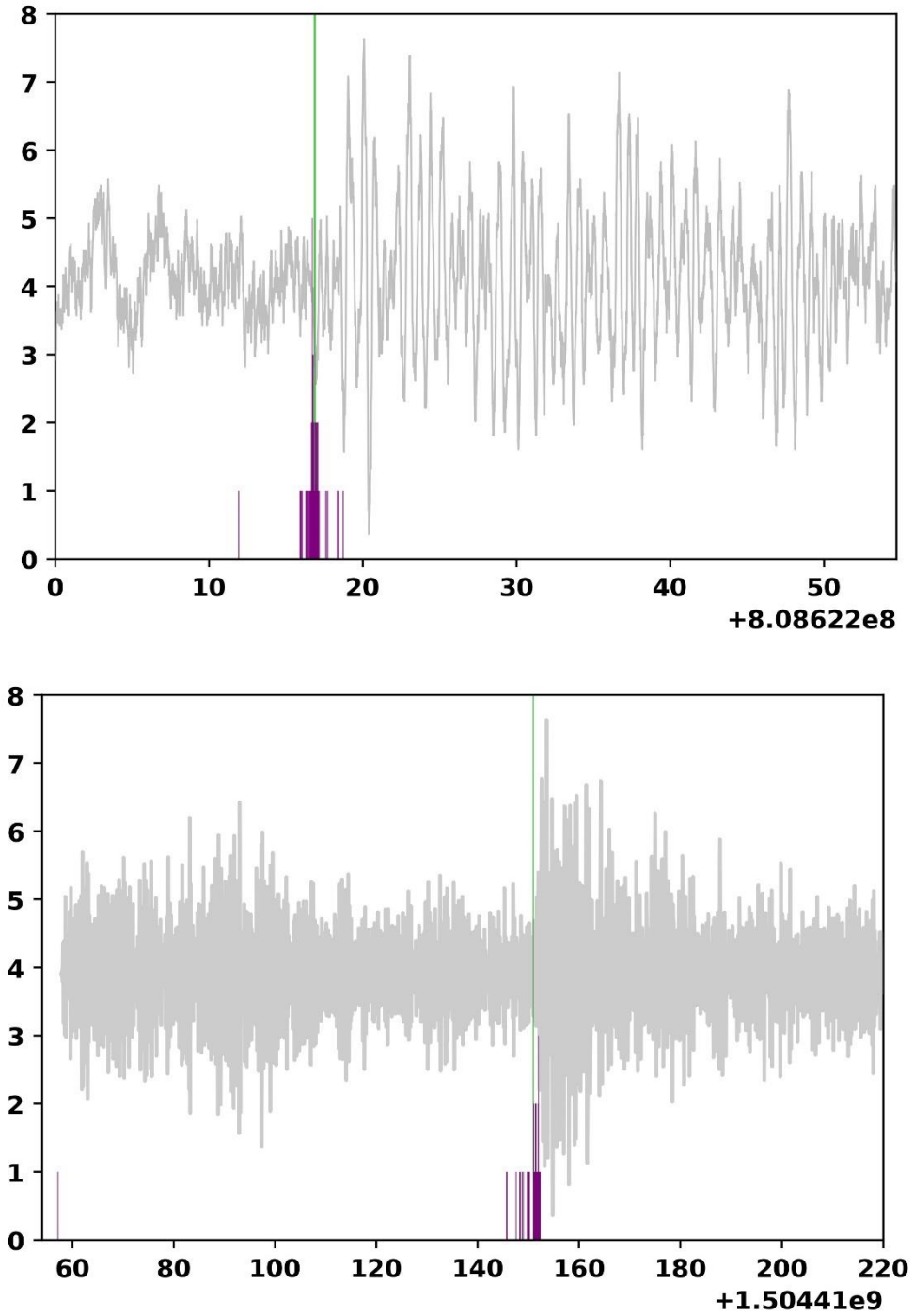

Figure A-1. Views of the entire waveforms that were used as the easier (top) and harder (bottom) within-subjects waveforms. The green lines mark the SME's pick for the best arrival time and the purple lines mark the participants' picks for the best arrival times across all visualization conditions.

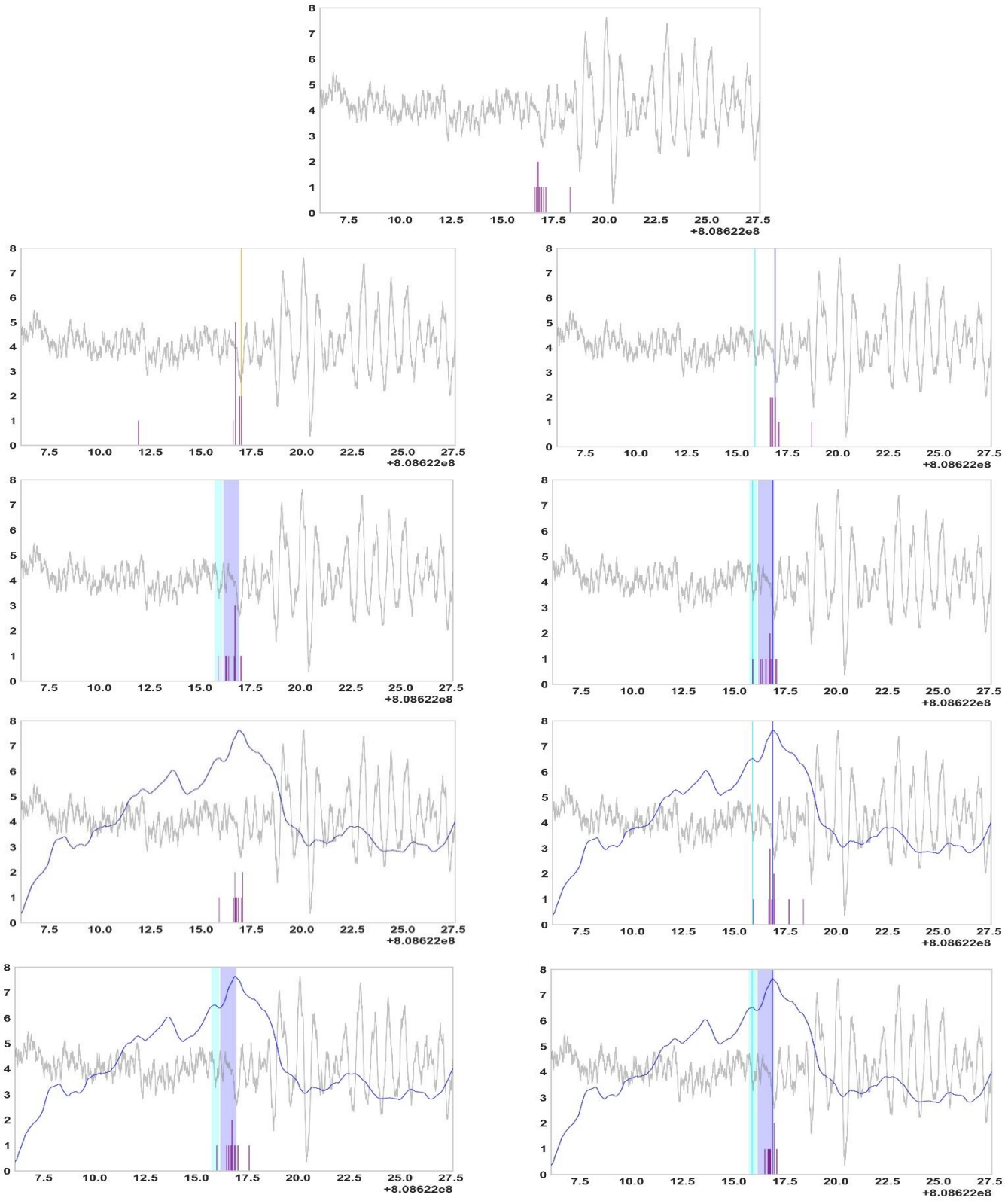

Figure A-2. The distribution of participants' best picks in each visualization condition for the easier within-subjects waveform.

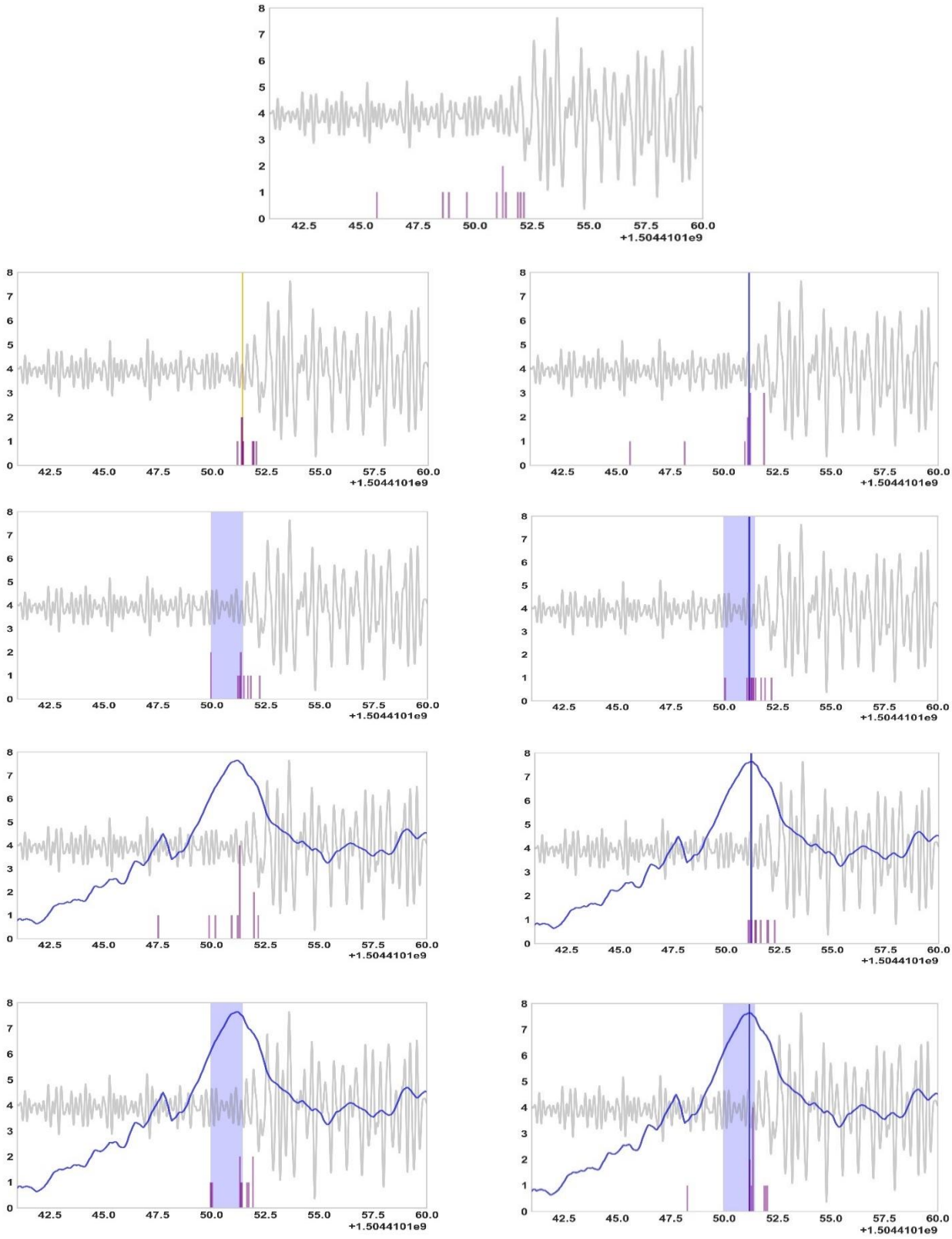

Figure A-3. The distribution of participants' best picks in each visualization condition for the more difficult within-subjects waveform.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Richard Field	05953	rvfield@sandia.gov
Travis Bauer	05954	tlbauer@sandia.gov
Christina Ting	05954	clting@sandia.gov
Jamie Coram	06371	jcoram@sandia.gov
Christopher Young	06371	cjyoung@sandia.gov
Susan Adams	06672	smsteve@sandia.gov
Michael Trumbo	06672	mctrumb@sandia.gov
Technical Library	01977	sanddocs@sandia.gov

Email—External (encrypt for OUO)

Name	Company Email Address	Company Name
Ronald Brogan	brogan.ronald@ensco.com	ENSCO, Inc.
James D. Morrow	jdmorrow@ieee.org	N/A

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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-NA0003525.