SAND2021-1850J

Generating Uncertainty Distributions for Seismic Signal Onset Times

Matt Peterson[*1], Charlie Vollmer[1], Ronald Brogan[2], David J. Stracuzzi[1], and Chistopher J. Young[1]

ABSTRACT

Signal arrival-time estimation plays a critical role in a variety of downstream seismic analyses, including location estimation and source characterization. Any arrival-time errors propagate through subsequent data-processing results. In this article, we detail a general framework for refining estimated seismic signal arrival times along with full estimation of their associated uncertainty. Using the standard short-term average/long-term average threshold algorithm to identify a search window, we demonstrate how to refine the pick estimate through two different approaches. In both cases, new waveform realizations are generated through bootstrap algorithms to produce full *a posteriori* estimates of uncertainty of onset arrival time of the seismic signal. The onset arrival uncertainty estimates provide additional data-derived information from the signal and have the potential to influence seismic analysis along several fronts.

KEY POINTS

- Can we estimate a full *a posteriori* uncertainty distribution over possible signal arrival times?
- Data-driven approaches can characterize the *a posteriori* uncertainty distribution of the signal arrival time.
- These uncertainty distributions can potentially influence seismic analysis along several fronts.

Supplemental Material

INTRODUCTION

Continued growth in seismic data collection necessitates improved automatic analysis methods to support data interpretation and decision making. One of the first analyses typically performed is onset (pick) determination. Seismic picking is usually done by a combination of automated analysis (see Withers *et al.*, 1998; Akram and Eaton, 2016, for reviews), which identifies candidate onset times, and human analysts, who refine the pick using domain expertise. The details of how this happens are important, because the precision with which the signal onset is identified can have a large impact on downstream analyses, such as event location.

Although experienced analysts provide the accepted gold standard for arrival-time picking, variation among analysts remains an issue. Zeiler and Velasco (2009) showed that not only do variations occur among analysts, but a single analyst will show variations when presented with the same data multiple times. In this article, we take the view that such variations amount to uncertainty among analysts in the true onset time of the signal. We show how to calculate the statistical uncertainty in the onset due to the data for any seismic waveform. We provide a framework that generalizes recent work in this domain (Simon *et al.*, 2020).

The goal of our data-driven approach is to estimate full *a posteriori* uncertainty distributions over possible signal arrival times. In this article, we consider bootstrap resampling methods for estimating the uncertainty distributions to avoid strong distributional assumptions. Importantly, the fully characterized pick-time distribution can increase the statistical rigor of downstream analyses by opening the door to uncertainty propagation throughout the data-processing sequence. The resulting posteriors can also be provided to analysts to aid in decision making.

In the remainder of the article, we first provide additional background on seismic signal detection and analysis and a review of existing data science methods. We then describe our framework for estimating the seismic arrival times and associated nonparametric uncertainty distributions. Next, we demonstrate the validity of our approach by applying two specific methods from the framework to a set of 95 waveforms and comparing the automated results to expert picks. Finally, we will conclude the article with a discussion of how to interpret the associated uncertainty distributions and thoughts on how they might be used.

1. Sandia National Laboratories, Albuquerque, New Mexico, U.S.A.; 2. ENSCO, Inc., Melbourne, Florida, U.S.A.

*Corresponding author: mgpeter@sandia.gov

Cite this article as Peterson, M., C. Vollmer, R. Brogan, D. J. Stracuzzi, and C. J. Young (2020). Generating Uncertainty Distributions for Seismic Signal Onset Times, *Bull. Seismol. Soc. Am.* **XX**, 1–10, doi: 10.1785/0120200125

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.

BACKGROUND

For many current seismic data-processing systems, signals in the processed data channels are detected using short-term average/long-term average (STA/LTA) energy detectors, which compute the ratio of energy in an STA to an LTA (Wahl, 1996a,b). A detection is declared when the STA/LTA ratio exceeds a predetermined threshold. After a detection is declared on a channel, a new detection cannot be made until the STA/LTA ratio falls below a predetermined value and a predetermined interdetection time has passed. The algorithm outputs onset times, which demarcate the transition between two distinct segments of a seismic waveform: a portion of background activity (often viewed as noise) and a portion with a distinguishable signal above the background.

The STA/LTA algorithm is designed for reliable signal detection, but only provides a crude estimate of the signal onset time. We propose a method for refining this crude estimate as well as providing a robust probability distribution around that estimate: a full measure of uncertainty.

The key to detecting a signal is identifying changes in frequency and/or amplitude of the time series when compared with persistent background. Ideally, a signal will have high signal-to-noise ratio (SNR) and a clear onset, in which case the uncertainty is insignificant. In particular, sharp, impulsive onsets provide a clear point of departure from background. In many cases, however, a nontrivial uncertainty is inevitable (Velasco *et al.*, 2001). Emergent signals in particular can have very large uncertainties, regardless of how large the SNR may eventually become. In these instances, an analyst may introduce filtering in an attempt to add clarity to the onset.

With emergent signals, we can make an initial pick and know that the signal is present after it. However, the true onset may occur earlier in time relative to our pick. Typically, the analyst will refine the onset time by considering the initial pick and shifting it earlier in the time series, looking carefully for changes in amplitude and/or frequency that could indicate the signal arrival. Recent work has been developed for both onset refinement and uncertainty estimation using Monte Carlo techniques. These researchers built confidence intervals for a wavelet-multiscale application of an Akaike information criterion (AIC)-based change-point detection scheme (Simon *et al.*, 2020). Our article generalizes their work and develops a complete framework to use onset refinement to build a full distribution of onset uncertainty by applying techniques well known in the machine learning and statistics communities.

Our approach to statistical refinement follows a similar strategy and then estimates the arrival-time uncertainty by applying techniques well known in the machine learning and statistics communities. One of the key ideas is bootstrapping (Efron, 1979), which permits a distribution-free (i.e., makes no assumption about the form of the underlying distribution) approach to assessing sampling variability through random sampling of data with replacement. Although many forms of bootstrap methodologies exist, we will focus on the parametric bootstrap (Hinkley, 1988; Efron and Tibshirani, 1993). Bootstrap applications have been widely studied and successfully applied over the past three decades in domains ranging from animal science (Casellas *et al.*, 2006) to genomics (Coombes and Biernacka, 2018).

Another technique commonly used in machine learning is adding random noise to data. This process perturbs the data to evaluate how robust an algorithm is to varying levels of noise. Image analysis (Van der Weken *et al.*, 2002), neural networks (Sietsma and Dow, 1991; Neelakantan *et al.*, 2015), and decision trees (Dietterich, 2000) have all used this type of approach to great effect. In this article, we will demonstrate both parametric bootstrap and noise-based methods to infer the uncertainty of an onset pick.

METHODOLOGY

Our goal is to estimate an *a posteriori* probability distribution over the signal arrival time for a given waveform. A viable approach should refine the preliminary detections provided by methods such as STA/LTA while remaining computationally efficient, so that it can be applied to large numbers of waveforms. In the following, we first outline a general approach to uncertainty estimation and then describe two instances of the general procedure in detail.

Starting with an unfiltered waveform from a single station, our method follows three steps. The first produces an initial onset time, k_0, along with a search window within the waveform for the refinement and uncertainty estimation methods. The initial onset can be selected using methods such as STA/LTA or manual picking. The search window, $W = \{s_1, \ldots, s_{k_0}, \ldots, s_n\}$, contains n data samples of the waveform centered around k_0 to provide sufficient data for estimating the statistical models used in subsequent steps while exploring the space of possible alternative onset times. In practice, the size of n depends on the sensor sampling rate, SNR, the signal emergence rate, and structure of the statistical models.

The second step uses the waveform data in the search window to refine k_0 by optimizing the fit of a statistical model to the data. Following Kamigaichi (1992), the refined onset-time estimate, $k_{0'}$, divides the search window into two regions: data points that precede $k_{0'}$ and those that come after. We assume that the first region, $\{s_1, \ldots, s_{k_{0'}}\}$, contains mostly background noise, whereas the second region, $\{s_{k_{0'}+1}, \ldots, s_n\}$, contains the seismic signal with a noise component. Importantly, there are a variety of choices for both the structure of the statistical models and the optimization criteria for finding the best $k_{0'}$ that splits the search window into noise and signal. In this article, we use Gaussian distributions to model the variation of the waveform's amplitude in the noise and signal regions, as well as consider two different optimization criteria as described in the following. Optionally, the data in the search window can be filtered prior to modeling, and our experiments show the impact of doing so.

Step three estimates a probability distribution over possible onset times by bootstrapping, which samples new waveforms, W_i, from the observed search window, W, and estimates an onset time for each sample. We assume that W is a random realization of the true, underlying seismic signal and that we can generate synthetic waveform realizations that maintain the same statistical properties as the observed data relative to our chosen statistical model. Thus, a Gaussian model will maintain the same mean and variance, whereas more sophisticated models may capture other properties of the observed waveform. As with the statistical modeling, there are several methods for generating a waveform and we describe two in the following. We use the same statistical model optimization technique used in step two to estimate an onset time, k_i, for each W_i. Collecting the k_i into a histogram produces an estimate of the true probability distribution over possible onset times.

The combination of methodological choices outlined previously for statistical models, waveform generation (sampling), and error (model fit) metric yields a family of possible methods for estimating the uncertainty in onset times. Although we do not explore this space here, different combinations of the three techniques will yield different results, and some combinations may be better suited to specific applications than others. Understanding the trade-offs involved remains a point of future work. The remainder of this section provides technical details for two instances of the general approach. We then validate and compare the performance properties of each instance in the Results and Discussion sections, respectively, to illustrate how the methodological choices combine to yield different performance properties.

Noise and signal models

The statistical models for both the noise and signal assume that the data were drawn independently and identically distributed from a Gaussian probability distribution with zero mean and finite variance. Let the observation at time t be $s_t \sim N(0, \sigma^2)$ in which $\sigma^2 < \infty$. To estimate the noise model, we simply calculate σ_n^2 for $\{s_1, \ldots, s_{k_0}\}$ in the selected search window. Similarly, we estimate the signal model, σ_s^2 for $\{s_{k_0}, \ldots, s_n\}$. The resulting models can be used both to estimate onset times (refinement) and to generate new waveforms as described in the following. Importantly, the quality of the refined onset time, the sampled waveforms, and resulting posteriors all depend on how well the statistical models capture the structural properties of W.

Many variants on this basic procedure are possible as described by Kamigaichi (1992). In practice, the noise and signal models need not assume the same structural properties. For example, autoregressive models may provide a better fit to the signal region of the data. See Kamigaichi (1994) for an example of using autoregressive models for onset determination, and see Brockwell and Davis (1991) or Box *et al.* (2015) for a comprehensive review of time-series methods.

Statistical onset-time estimation

Statistical pick refinement is based on two choices: the statistical model forms used to represent the data and the optimization criterion used to measure how well the models fit the data. We model our data with the Gaussian models described previously for simplicity. To demonstrate the optimization criteria, we choose the commonly used AIC (Akaike, 1974) along with the chi-squared distance (Greenwood and Nikulin, 1996) as an alternative. Both methods begin by estimating noise and signal models and then refine k_0 by minimizing the distance between the models and the waveform, W. The resulting minima indicate the point at which W transitions from the noise model to the signal model. The different criteria yield slight differences in the optimization methods, which we detail next.

The AIC-based approach optimizes the pick directly by manipulating $k_{0'}$ to minimize the complexity and maximize likelihood of the combined noise and signal models. For each possible index k in W, we estimate σ_n^2 and σ_s^2 as earlier and then calculate the familiar AIC metric:

$$\mathrm{AIC}(k) = k \ln(\sigma_n^2) + (n - k) \ln(\sigma_s^2). \quad (1)$$

The value of k that minimizes the metric (the noise and signal models that lose the least information) identifies the refined onset time. Kamigaichi (1992) describes a number of variations on the basic AIC approach described here, the details of which are not important for our purposes of quantifying uncertainty in onset times. The center panel of the AIC result in Figure 1 shows an example of the AIC score curve and its minimum for an example waveform.

The chi-squared approach optimizes the pick by estimating a local fit metric and then finding the point that minimizes the metric. The method begins by calculating distribution models of the amplitude within the noise and signal parts of the window as described previously. In this case, we leave a buffer around k_0, using the early third of W to estimate the noise amplitude distribution $N(0, \sigma_n)$ and the late third to estimate $N(0, \sigma_s)$; these are our reference models for noise and signal. For simplicity, we assume that the signal persists throughout the late third of W. More complex scenarios, such as a short, impulsive signal, may require modified strategies, which we do not consider here. Given these two models, we next generate a score curve to indicate the transition from noise model to signal by sliding a static-sized subwindow, $V = \{s_p, \ldots, s_q\}$ such that $1 \le p, q \le n$ and $q - p \ll n$, across W and calculating the chi-squared distance between the amplitudes in V and the reference noise and signal models. We then sum these two distance scores for each observation in W and select the minimum value for the refined pick, $k_{0'}$.

Chi-squared distance operates on histograms, so we use the empirical histogram values to calculate the distance rather than the estimated σ^2 parameter values. Let g and h represent the amplitude histograms for W and V, respectively, in which g

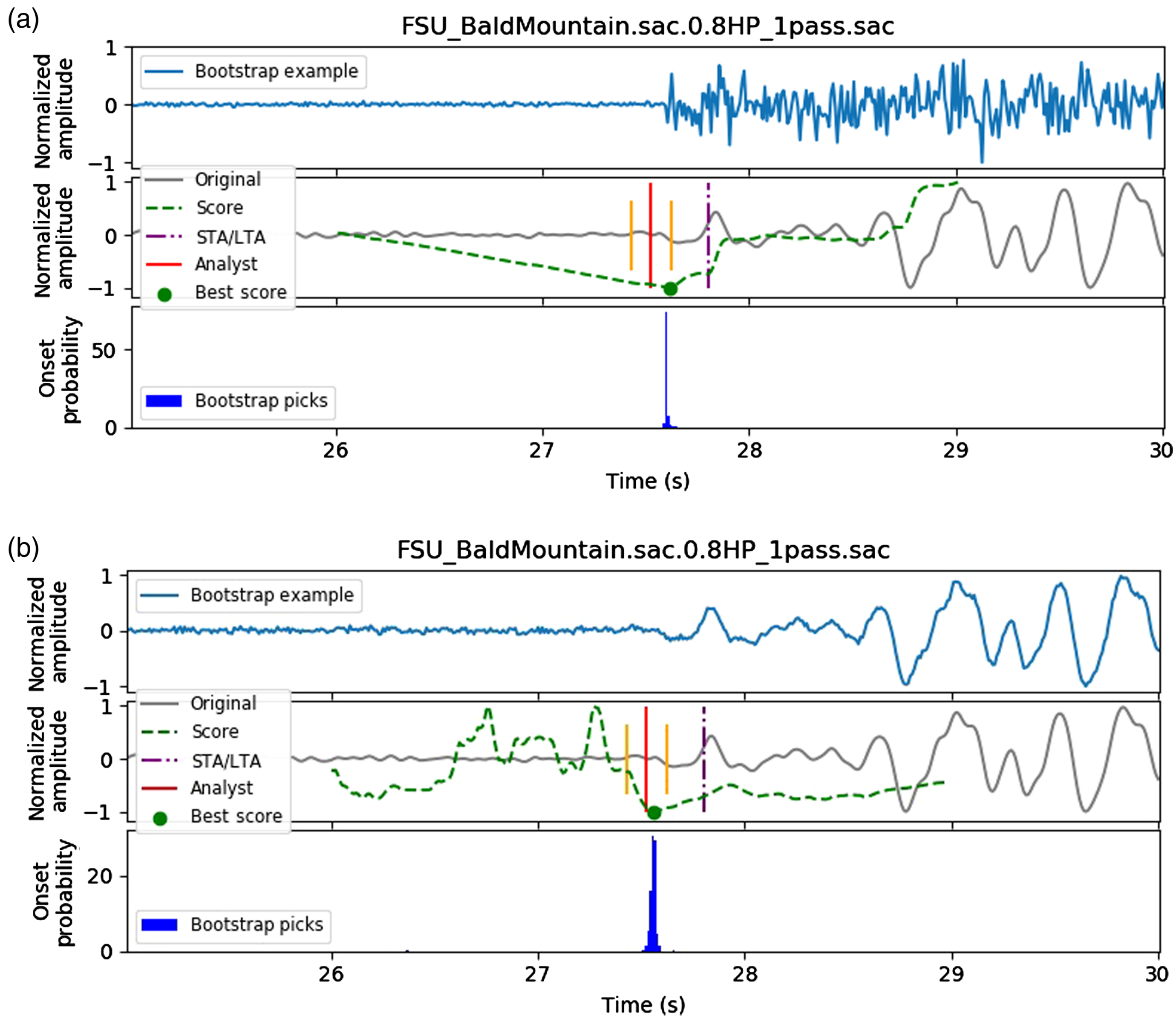

Figure 1. Example results for both uncertainty estimation methods on the same waveform: (a) Akaike information criterion (AIC)-based scoring with noise and signal sampling and (b) chi-squared scoring with noise sampling only. The top plot in each result shows a sampled waveform. The center plot shows the original waveform with the score curve, minimum score, the analyst's most likely pick (long solid vertical bar), the analyst's early and late plausible picks (short solid bars), and the short-term average/long-term average (STA/LTA) pick (dashed bar). The bottom plot shows the posterior probability distribution over onset times. The color version of this figure is available only in the electronic edition.

can represent either the noise or signal model, respectively. Then, the chi-squared distance is calculated as

$$d(g,h) = \sum_i \begin{cases} \infty & \text{if } g_i = 0 \text{ and } h_i > 0 \\ 0 & \text{if } g_i = 0 \text{ and } h_i = 0 \\ \frac{(h_i - g_i)^2}{g_i} & \text{otherwise} \end{cases}, \quad (2)$$

for $1 \le i \le |h|$. The resulting distance corresponds to the sample at the midpoint of V. For the experiments detailed in the following, we used $|V| = q - p = 0.3$ s (the number of samples depends on the sensor sampling rate) or 10% of our search window, and $|h| = 1001$ histogram bins uniformly distributed between the minimum and maximum amplitudes observed in W. We have not attempted to optimize these parameters. Figure 1b shows the chi-squared score curve and its minimum for the same example waveform used with AIC. The difference can be noted in the score curve shapes for the two methods.

Bootstrapping the onset-time uncertainty

The goal of bootstrapping in our method is to estimate a probability distribution over possible onset times for the original search window, W. Earlier, we assumed that W was a random realization of the true seismic signal. Bootstrapping generates many more random realizations, W_i, for which we then estimate onset times k_i as described previously for W. The resulting distribution over k_i indicates the probability that the true, unobservable signal started at a given time. Importantly, both of the sampling methods presented in the following are based on the statistical models used to model W. The accuracy of the estimated uncertainty distribution therefore depends on the degree to which these models capture the relevant statistical properties of W.

We describe two approaches to sampling new W_i from W, both of which use models of W (parametric bootstrap). The first involves adding noise to W using the noise model calculated for our initial onset refinement. For each sample s_i in W, we randomly draw an amplitude from $N(0, \sigma_n^2)$ and add it to s_i (always starting from the original W). The resulting W_i has the same underlying signal, but a different (statistically equivalent) noise component. The top plot of the chi-squared example in Figure 1 shows a W_i constructed for the W shown in the middle plot, whereas the lower plot shows the resulting posterior probability distribution as a histogram. The mode of the distribution (peak of the posterior) indicates the most likely onset time.

The second approach constructs the W_i by sampling from both the noise and signal models for W. In this case, we draw random samples from $N(0, \sigma_n^2)$ for $s_1, \ldots, s_{k_0'}$ of W_i and from $N(0, \sigma_s^2)$ for $s_{k_0'+1}, \ldots, s_n$. The resulting waveform may appear substantially different from the observed waveform, but its statistical properties should be equivalent relative to the models σ_n^2 and σ_s^2. The top plot of the AIC example in Figure 1 shows a waveform drawn using the second approach, with the associated posterior in the lower panel. Again, notice the difference in both the sampled waveforms and the resulting posteriors. These follow from the differences in the model optimization metrics and the sampling methods. Changes to the statistical model form would further alter the resulting metric scores, sampled waveforms, and posteriors.

Summary

At the beginning of the Methodology section, we stated that the combination of methodological choices outlined earlier for statistical models, waveform generation (sampling), and error (model fit) metric yields a family of possible methods for estimating the uncertainty in onset times. We reiterate this point here to clarify the results and discussion that follow. The two metrics described (AIC and chi-squared) could easily be crossed with the two sampling techniques, which implies a 2×2 design. However, the design space is more accurately described as m modeling methods crossed with n metrics crossed with p sampling techniques. We chose two relatively simple approaches to demonstrate, but the full space of possibilities requires exploration. The Discussion section identifies some practical implications of the choices we made and provides direction on how to navigate the design space.

RESULTS

Our goal is to estimate the arrival-time uncertainty as a probability distribution in a statistically rigorous manner. To validate our methods, we would ideally demonstrate two properties of our distributions. First, the most likely onset times identified by our posterior estimates (distribution mode) should be consistent with an expert analyst's picks. Experts can reasonably disagree over the exact onset, but our statistical best guess should be similar to that of an expert. Second, our posterior estimates should reflect the uncertainty in the underlying waveform. In practice, no ground truth for our probability distributions exists, so we must rely on rough estimates by an expert seismic analyst.

An expert seismic analyst selected 95 waveforms from the Incorporated Research Institutions for Seismology (IRIS) Data Management Center using a variety of stations, source distances, signal qualities, and ranging from emergent to impulsive onsets. The signals cover a range of events including earthquakes and mining explosions. The analyst picked the first arriving P phase for each waveform six times: unfiltered, with a 0.8 Hz high-pass filter, and with 1–3, 2–4, 3–6, and 4–8 Hz band-pass filters. In general, the analyst first attempted to pick on the unfiltered waveforms, then the 0.8 Hz high-pass, and then the band-pass filters. The analyst also indicated which of the six variants provided the preferred (best) pick determination. In many instances, adjacent bands offer similar representations, so SNR was a consideration.

For all six versions of every waveform, the analyst provided three estimates. First, the most likely onset time corresponds to the analyst's best guess of onset time. The remaining two estimates correspond to the earliest and latest plausible times that the analyst believed the onset could have occurred. In a few cases, some of the filters did not provide sufficient information for the analyst to make a pick, so those were skipped over. Although we can reasonably compare the mode of our posteriors with the analyst's most likely onset time, we cannot directly compare the distribution width or confidence intervals to the analyst's plausible time interval because we cannot determine if the plausibility bounds correspond to a consistent probability level. We use the analyst's plausible interval to get an approximate sense of whether our distribution is reasonable and to highlight the challenge of validating an uncertainty distribution on observational data.

To compare our method's most likely onset times relative to the analyst's, we used all five filtered versions of the 95 waveforms. For consistency across examples and to ensure sufficient data for our statistical methods, we extracted a 9 s search window centered around the analyst's best picks and we did not attempt to optimize the window size. We applied both statistical methods (AIC score metric with signal and noise sampling, and chi-squared metric with noise sampling only) and STA/LTA (threshold 1.5, 2 s STA, 4 s LTA) to the windows. Detriggering was not a concern as we only analyzed the

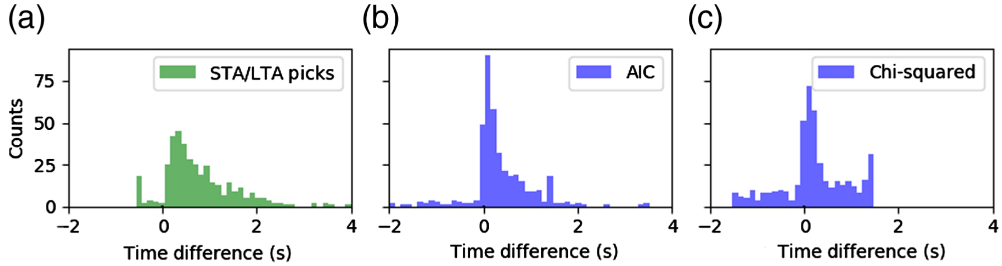

Figure 2. A comparison of the analyst's most likely onset times to (a) STA/LTA and the most likely onset time identified by the (b) AIC and (c) chi-squared methods. A time difference of zero means the analyst and computed picks agree, negative values indicate that the computed pick was early, and positive values indicate the computed pick was late relative to the analyst. The color version of this figure is available only in the electronic edition.

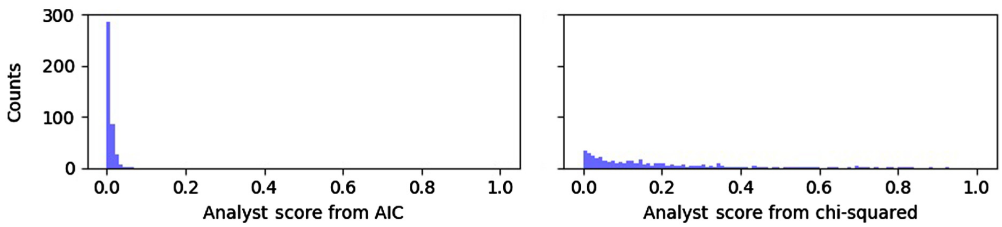

Figure 3. Histograms of the optimization scores assigned to the analyst's most likely onset time. Zero indicates that the analyst's onset time corresponds to the minimum of the score curve. The color version of this figure is available only in the electronic edition.

first-arrival *P* wave in our experiments. The search windows are not generally equivalent across all five filters, as the analyst did not pick the same onset time for each.

Figure 2 shows the difference in seconds between the analyst's picks and the computed picks. A value of zero means that the analyst's pick aligns perfectly with the computed onset time, whereas positive and negative values correspond to late- or early-computed onset times, respectively. The results show that STA/LTA and AIC tend to pick late relative to the analyst, although AIC tends to pick much closer to the analyst than STA/LTA. The chi-squared method onset-time estimates tend to be more evenly distributed around zero. Complete results for all 95 waveforms are available in the supplemental material available to this article.

To understand the relationship between the optimization metrics and the analyst's picks, we also plotted a histogram of the scores calculated at the analyst's most likely onset time. To normalize across waveforms, we calculated the score relative to the range of scores observed in the search window. A score of zero at the analyst's chosen onset time corresponds to the minimum of the score curve, or perfect agreement. A score of one corresponds to the maximum of the score curve. The score curve generally assigns low (good) scores to the analyst pick, meaning that even if they do not agree on the exact onset time, the assigned score values agree with the analyst to a substantial degree. Figure 3 shows that the AIC metric tends to agree well with the analyst, generally within 5% of the minimum. The chi-squared metric shows more variance in the scores assigned to the analyst's onset time, although still usually within 20% of the minimum score.

DISCUSSION

The aggregated results of the previous section suggest that our methods are able to appropriately model the statistical properties found in both the noise and signal aspects of the waveforms and refine the onset arrival time in accordance with a trained expert seismic analyst. Moreover, this methodological approach is able to capture more information in the signal than simply the onset time. The correctness of the optimization score curves and the posterior probability distributions is more difficult to validate. However, we can attempt to understand the properties of the scores and posteriors by examining a few examples. We hypothesize that the information contained in the scores and posteriors may be informative to an analyst.

To further investigate the score curves, consider the results shown in Figure 4, which shows the results of both the AIC and chi-squared approaches for five different filters on a single waveform. The score curves (dashed curve), like the posterior probability distributions, can be viewed as a measure of statistical evidence that the onset occurs at a given time. The score curves are calculated with respect to the original, filtered waveform only, not the sampled waveforms. Lower scores indicate more evidence of an onset relative to higher scores. For example, the 4–8 Hz filter for AIC shows a large section of relatively constant scores between 21 and 23.5 s, which indicates that the statistical model views the entire range as showing similar evidence of onset. In contrast, the 3–6 Hz filter for AIC shows two local minima with similar scores, which implies that the statistical model has identified two distinct candidate onset times. Finally, the 0.8 Hz high-pass filter for AIC shows a single minimum corresponding to the analyst pick. In this case, the statistical approach disagrees with the analyst's pick for the filter

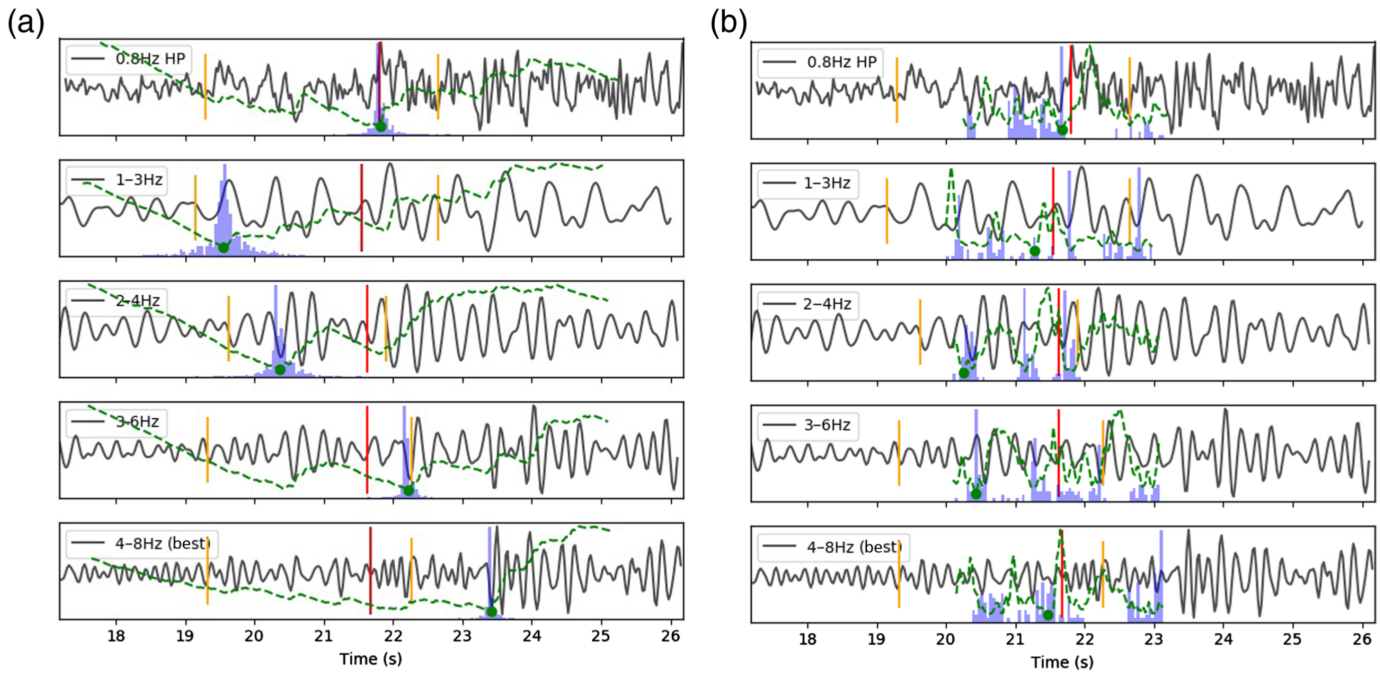

Figure 4. Results produced by the (a) AIC and (b) chi-squared approaches on five different filters for a single waveform. Plots show the score curve (dashed), posterior probability distribution (histogram), analyst earliest and latest (short vertical bars), and most likely onset times (long vertical). HP, High-pass Filter. The color version of this figure is available only in the electronic edition.

identified by the analyst as best (4–8 Hz), but they agree on the 0.8 Hz high-pass filter.

In contrast, all of the score curves for the chi-squared approach in Figure 4 show multiple minima spread across the search window. For chi-squared, the score curves are only calculated for a 3 s search window around the initial onset determination (here, the analyst pick). The score curves have the same interpretation as earlier: the minima indicate evidence of a transition, and large (or multiple) areas of near-minimal scores indicate a degree of confusion. In this case, the score curve for all five filters suggests that the signal onset is not well defined and that there are several candidate transitions. Importantly, this conclusion is consistent with the relatively wide range of plausible onset times indicated by the analyst.

Turning now to the posterior probability distributions (histograms) in Figure 4, the most salient feature of the AIC method with signal and noise sampling is that they tend to distribute approximately normally around the score curve minimum. This property follows from the sampling method of drawing observations from the noise model up to the score minimum, and then drawing from the signal model, which necessarily places the transition in the sampled waveforms at a fixed point in time. The width of the resulting posterior then indicates how well the optimization metric, AIC in this case, can identify the transition point for a variety of similar waveforms. In general, a narrow distribution implies good identification whereas a wide distribution implies poor identification. Thus, the posterior width is proportional to the SNR.

Taken in conjunction with the previous discussion about the shape of the score curve, we can conclude that for difficult cases like the 4–8Hz filter, the large flat region of the AIC score curve provides more information about the uncertainty of the onset time than the actual AIC-based posterior, which is narrow, but poorly located. In the 0.8 Hz high-pass filter case, however, the AIC-based posterior is highly informative because there is a unique minimum on the score curve and the width of the distribution indicates how well the statistical model can identify the transition from noise to signal. This raises an interesting question for future work: can we use the combination of score curve and posterior to automate identification of an ideal filter for statistical analysis of arbitrary event waveforms? This is an important issue, as we have already seen evidence that an analyst's choice of filter may not correspond to the ideal filter for computational methods.

The posteriors for the chi-squared method with only noise sampling are consistent with the score curve. The score minima correspond to areas of higher mass in the probability distribution. The relatively poor quality of the signal in this example yields a broadly distributed and multimodal histogram. The next example shows a more viable application of the chi-squared with noise-only sampling approach.

Figure 5 shows the results produced by both methods on a second waveform. The results for AIC with signal and noise sampling are qualitatively similar to the previous example. In this case, the filter identified as best by the analyst (4–8 Hz) also produces the narrowest probability distribution. The

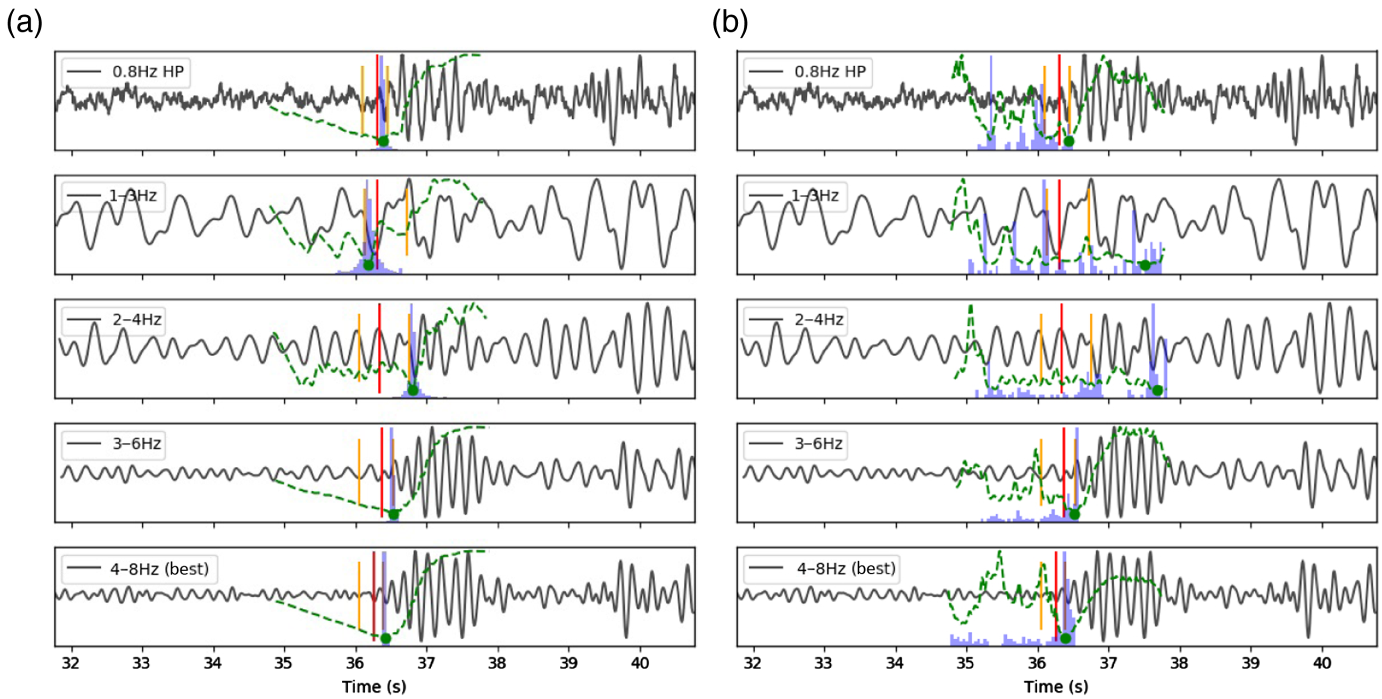

Figure 5. A second example of results produced by the (a) AIC and (b) chi-squared approaches on five different filters for a single waveform. Plots show the score curve (dashed), posterior probability distribution (histogram), analyst earliest and latest (short vertical bars), and most likely onset times (long vertical). The color version of this figure is available only in the electronic edition.

higher SNR of the waveform in Figure 5 also produces narrower distributions across all of the different filters.

The chi-squared score curves in Figure 5 show more varied behavior in this example. For example, some curves have flat bottoms (1–3 and 2–4 Hz), multiple distinct local minima (0.8 Hz), and clear global minima (3–6 and 4–8 Hz). However, the noise-only sampling approach used for the chi-squared method yields important differences in the posteriors. The 0.8 Hz high-pass filter provides an excellent example. The score curve minimum appears at approximately 36.5 s and corresponds to one mode of the posterior probability distribution. However, unlike the signal and noise sampling approach used for AIC, the noise-only sampling approach used for chi-squared also produces a mode near 36.0 s that corresponds to a local minima in the score curve, and another mode near 35.3 s that corresponds to a third local minima. In this case, the earliest of the modes has the highest probability, suggesting that the statistically most likely onset time (given a 0.8 Hz high-pass filter) is near 35.3 s.

Although this does not agree with the analyst's assessment, the result is important for two reasons. First, the presence of three modes suggests three specific onset-time candidates that an analyst might consider. Second, the relatively low probabilities associated with each of the modes suggest that the automated picker is not confident in any of the candidate onset times, which suggests that either the filter is inappropriate to the given waveform, or that there is simply poor statistical evidence for any specific onset time.

In the 4–8 Hz case, the bulk of the posterior distribution is located within a small time frame, which indicates clear support for picking in that region. Given a posterior in which most of the mass is concentrated in one region, we may be able to infer that the width of the region indicates how well the statistical model captures the onset as discussed previously. However, it is not clear that this will always be true for the noise-only sampling method. The noise-based approach adds noise to the original waveform, thereby retaining whatever transition was present in the original. As a result, the estimated posterior better reflects the probability that the transition may occur at an arbitrary location in the waveform (as opposed to one location hypothesized by the signal and noise approach) at the expense of a clear relationship between posterior width and SNR. The difference raises a second important question for future work: what should the posterior probability distribution represent? The answer likely depends on how the posterior will be used, for example, to improve onset-time determination versus propagation through location analysis, but may reveal a variety of capabilities not considered here.

The examples shown in Figures 4 and 5 also raise important questions about the scope of uncertainty analysis for onset-time determination. The modeling and sampling methods described previously capture properties of the noise and signal in the waveform, which can be broadly described as estimating the uncertainty due to the data. However, other sources of uncertainty exist. For example, the onset-time probability distributions shown in Figures 4 and 5 often show limited overlap across the different filters. Each possible filter changes the data, and so yields different results. We cannot say for sure which filter is best, so this becomes another source of uncertainty—uncertainty due to the data preprocessing. In theory, we could merge the posteriors due to the various filters into a single probability distribution that captures uncertainty due to the data and the preprocessing. In practice, some filters are more appropriate than others, so we need a way to weight the different filter results, which remains a point of future work.

Other sources of uncertainty include the statistical model. As noted in the Methodology section, time-series methods such as autoregressive models will capture different aspects of the waveform compared to the Gaussian models used earlier. However, several parameters govern the specific structure of autoregressive models, and as with the choice of filter, the "correct" parameter settings can be difficult to determine. Thus, questions about which parameters to use and how to combine the different posterior probability distributions that they produce mirror questions about the filtering process and yield what we can term model form uncertainty.

Estimating and combining all sources of uncertainty remains a point of future research. In practice, this means that our uncertainty analysis is incomplete. Nevertheless, the results and discussion suggest that simply estimating the uncertainty due to the data alone, particularly in conjunction with the score curves, can provide rich and nuanced information about the statistical properties of waveforms that is not otherwise available.

CONCLUSIONS

In this article, we described a general methodology for estimating the uncertainty in seismic onset-time determination. Our approach provides a posterior probability distribution over possible onset times, and we have validated its efficacy using a selection of examples from the IRIS Data Management Center manually labeled by an expert seismic analyst. Our work provides two key lessons. First is that our methodology outlines a family of possible pick refinement and uncertainty estimation techniques that can be tailored to specific needs by manipulating the form of the underlying statistical models, the optimization criteria, and the bootstrap-sampling methods. Our results provide a proof of concept that onset-time uncertainty can be reliably estimated from the waveforms alone, and additional experimentation with the technical details may yield substantial performance improvements.

The second lesson of our work is that the posterior probability distributions and underlying model fit scores may provide a wealth of statistical insight to both seismologists and the downstream computational analyses that depend on onset-time detection. All statistical models are limited by their assumptions, but we have shown that even simple models can provide evidence about the likely timing of signal onset that sometimes differs from analyst intuition. Such information could be used to support waveform filter selection or to develop a more collaborative analysis environment in which expert knowledge fluidly combines with rigorous statistical analytics. At a minimum, propagating the posterior probability distributions to downstream analyses, such as location determination, may increase their fidelity, particularly when the estimated posteriors have heavy tails or multiple modes. Each of these applications requires further study and experimentation. Our work shows that the rich information required to support these applications is computable.

DATA AND RESOURCES

Seismograms used in this study were downloaded from the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (https://ds.iris.edu/ds/nodes/dmc/; last accessed October 2019). Our supplemental material contains additional details about the specific stations used, as well as the complete list of results for both the Akaike information criterion (AIC)-based and the chi-squared onset refinement and uncertainty estimation algorithms described in the main article.

ACKNOWLEDGMENTS

The authors thank Nicole McMahon and Dylan Anderson for many helpful comments on earlier drafts of this work. The authors acknowledge the support of the National Nuclear Security Administration Office of Defense Nuclear Nonproliferation Research and Development for funding this work. This article describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the U.S. Government. 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. This research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D). The authors acknowledge important interdisciplinary collaboration with scientists and engineers from Los Alamos National Laboratories (LANL), Lawrence Livermore National Laboratories (LLNL), Mission Support and Test Services (MSTS), Pacific Northwest National Laboratories (PNNL), and Sandia National Laboratories (SNL).

REFERENCES

Akaike, H. (1974). A new look at statistical model identification, *IEEE Trans. Automat. Contr.* **19,** no. 6, 716–723.

Akram, J., and D. W. Eaton (2016). A review and appraisal of arrival-time picking methods for downhole microseismic data, *Geophysics* **81,** no. 2, KS71–KS91.

Box, G., G. Jenkins, G. Reinsel, and G. Ljung (2015). *Time Series Analysis: Forecasting and Control*, Fifth Ed., John Wiley & Sons, Hoboken, New Jersey.

Brockwell, P. J., and R. A. Davis (1991). *Time Series:Theory and Methods*, Springer, New York, New York.

Casellas, J., J. Tarres, J. Piedrafita, and L. Varona (2006). Parametric bootstrap for testing model fitting in the proportional hazards framework: An application to the survival analysis of Bruna dels Pirineus beef calves, *J. Anim. Sci.* **84,** no. 10, 2609–2616.

Coombes, B. J., and J. M. Biernacka (2018). Application of the parametric bootstrap for gene-set analysis of gene-environment interactions, *Eur. J. Hum. Genet.* **36,** no. 11, 1679–1686.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization, *Mach. Learn.* **40,** no. 2, 139–157.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Ann. Stat.* **7,** no. 1, 1–26.

Efron, B., and R. Tibshirani (1993). *An Introduction to the Bootstrap*, Chapman and Hall/CRC, London, United Kingdom/Boca Raton, Florida.

Greenwood, P. E., and M. S. Nikulin (1996). *A Guide to Chi-Squared Testing*, Vol. 280, John Wiley & Sons, Hoboken, New Jersey.

Hinkley, D. (1988). Bootstrap methods, *J. Roy. Stat. Soc.* **50,** no. 3, 321–337.

Kamigaichi, O. (1992). A fully automated method for determining the arrival times of seismic waves and its application to an online processing system, *Proc. of the 34th GSE Session, GSE/RF/62*, Geneva, Italy, 4–6 November.

Kamigaichi, O. (1994). Automated identification of arrival time, etc., using AR-model, *Proc. of the GSE Workshop*, Tokyo, Japan, 20–22 June.

Neelakantan, A., L. Vilnis, Q. V. Le, I. Sutskever, L. Kaiser, K. Kurach, and J. Martens (2015). Adding gradient noise improves learning for very deep networks, available at https://arxiv.org/abs/1511.06807 (last accessed October 2020).

Sietsma, J., and R. J. Dow (1991). Creating artificial neural networks that generalize, *Neural Networks* **4,** no. 1, 67–79.

Simon, J. D., F. J. Simons, and G. Nolet (2020). Multiscale estimation of event arrival times and their uncertainties in hydroacoustic records from autonomous oceanic floats, *Bull. Seismol. Soc. Am.* **3,** no. 110, 970–997.

Van der Weken, D., M. Nachtegael, and E. E. Kerre (2002). An overview of similarity measures for images, *2002 IEEE International Conf. on Acoustics, Speech, and Signal Processing*, Vol. 4, IEEE, IV–3317.

Velasco, A., C. Young, and D. Anderson (2001). *Uncertainty in phase arrival time picks for regional seismic events: An experimental design*, *Tech. Rept.*, U.S. Department of Energy, Washington, D.C.

Wahl, D. (1996a). Programmer's guide for the detection and feature extraction program, *Tech. Rept. SAIC-96/1069*, Science Applications International Corporation, Reston, Virginia.

Wahl, D. (1996b). User's manual for the detection and feature extraction program, *Tech. Rept. SAIC-96/1098*, Science Applications International Corporation, Reston, Virginia.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. Seismol. Soc. Am.* **88,** no. 1, 95–106.

Zeiler, C., and A. A. Velasco (2009). Seismogram picking error from analyst review (SPEAR): Single-analyst and institution analysis, *Bull. Seismol. Soc. Am.* **99,** no. 5, 2759–2770.

Manuscript received 20 March 2020
Published online 3 November 2020