SAND2018-3624J

Implementation of the Waveform Correlation Event Detection System (WCEDS) method for regional seismic event detection in Utah

Stephen Arrowsmith, Christopher Young, Kristine Pankow

Abstract

This paper extends previous work on the development of a backprojection method for local and regional seismic event detection, the Waveform Correlation Event Detection System (WCEDS). The improvements outlined here make the technique much more flexible for local and regional earthquake or explosion monitoring. We first explore how WCEDS can be formulated using either a synthetic or empirical approach, showing that the synthetic approach is much more flexible and can lead to the detection of smaller events, and to significant improvements in the resolution of event parameters. Second, we discuss the factors that influence the grid of event hypotheses used for backprojection, and develop an algorithm for generating suitable grids on the basis of the network resolution. Third, we explore the effect of including different phases in the backprojection operator, showing that the best results for the study region can be obtained using only the Pg phase, and by including terms for penalizing early arrivals when evaluating the fit for a given event hypothesis. Fourth, we incorporate two parallel backprojection computations with different distance thresholds to enable the robust detection of both network-wide and small (sub-network-only) events. The set of improvements are outlined by applying WCEDS to four example events on the University of Utah Seismograph Stations (UUSS) network.

1. Introduction

Most seismic event detection methods are based on a standard paradigm for processing seismic data, which involves separate station-level and network-level processing tasks (Le Bras and Wuster 2002). An alternative class of methods, which we refer to here as back-projection methods, treat the two processing levels in a single step. Back-projection methods involve a grid-search over possible event hypotheses where, for every hypothesis, the consistency of the observed waveform data with an event at that hypothesis is evaluated. To evaluate the consistency, several different methods can be used. As some examples, Shearer (1994) and Arrowsmith et al. (2016) use a matched filter to correlate time-and-distance aligned STA/LTA processed data, for a given event hypothesis, with an empirical stack. As a contrasting approach, Withers et al. (1999), Baker et al. (2005) and Kao and Shan (2004) calculated the total power in the waveforms across a network in model-predicted time windows given an event hypothesis.

The Waveform Correlation Event Detection System (WCEDS) method is a generalized back-projection method that can either be based on an empirical matched filter approach (Arrowsmith, et al. 2016), or on summing the power in predicted time windows across a network (Withers, et al. 1999, Young, et al. 1998, Young, et al. 1996, Young, et al. 1996). To avoid confusion, it is worth noting that the use of the term 'waveform correlation' in the acronym does not imply correlation in the mathematical sense; it implies some measure of the correspondence of a property of the observed waves recorded across a seismic network with some expected property for a given event hypothesis (as described in Equation 1 below, 'correlation' is analogous to the stack power). If we consider that the seismogram at the ith station in a network, $x_i(t)$, is preprocessed in some way such that $a_i(t) = \gamma[x_i(t)]$, where $\gamma[\dots]$ represents the combination of a bandpass filter followed by a STA/LTA transform, and that the travel-time prediction for a source at point $\underline{r} = (\phi, \lambda)$ to station i is $t_{\underline{r}i}$, then the stack power for a single phase at $\underline{r}$ for an event at origin time τ can be defined as:

$$P(\tau, \underline{r}) = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{m=0}^{M-1} a(\tau + t_{\underline{r}i} + m\delta t)\right]. \quad (1)$$

In this equation, N is the number of stations, τ is a trial origin-time, M is the number of points within a time window that starts at the predicted arrival time, and δt is the sampling interval. Most travel-time based back-projection methods are based on a version of Equation (1). A common term for such methods is Reverse Time Migration (RTM). Most RTM methods calculate P over a grid of possible event hypotheses using a version of Equation 1, and then find events by searching for peaks in location and time. If the travel-time prediction varies only with distance, Young et al. (1996) showed that Equation 1 can be calculated more efficiently by a dot product formulation that calculates all possible combinations of each observed waveform with the full set of predictions by distance, then sums these as appropriate for each event hypothesis location. In this study, the dot product formulation is implemented as follows:

$$\underline{X} = \left(\underline{M} \cdot \underline{D}\right)^T$$
$$P(\tau, \underline{r}) = \frac{1}{N_\Delta}\sum_{i=1}^{N_\Delta} X_{i\underline{r}} \quad (2)$$

where $\underline{M}$ is an 'image matrix' of dimensions $N_D \times N_T$, $\underline{D}$ is a time-window of the network data, arranged in a matrix of dimensions $N_T \times N$, N_D is the number of distance bins, N_T is the number of time bins, $X_{i\underline{r}}$ is the element in $\underline{X}$ that corresponds to a given station and distance (these are stored as a lookup table for each grid node and station), and N_Δ is the number of stations within some threshold distance, Δ. The image matrix, $\underline{M}$, can be formulated with a boxcar function of width M and amplitude 1, starting at the predicted travel time for a given distance. Equation 1 is the more general form of Equation 2 that does not assume only distance dependence (i.e., Equation 1 can be applied for 3D Earth models, but Equation 1 assumes a 1D Earth model). However, Equation 2 can also be implemented using a stack of azimuthally-averaged data to generate $\underline{M}$, as was done by Shearer (1994) and Arrowsmith et al. (2016).

Arrowsmith et al. (2016) outlined an implementation of WCEDS for continuous data processing using empirical data to build the image matrix, illustrating the results on one-day of data from the University of Utah Seismograph Stations (UUSS) network. The one-day dataset contained a number of small events recorded by nearby dense clusters of stations in Northern Utah. This study extends and improves upon the method presented in Arrowsmith et al. (2016) in a number of ways, which collectively make the algorithm more adaptible for broad-area monitoring. Key improvements include the use of synthetic, as well as empirical image matrices; the use of different distance thresholds to enable robust detection of both large and small events; and the use of a spatially-variable grid of event hypotheses to improve the resolution of events near relatively dense areas of the network and mitigate against forming false events using phases outside the network. The new improved version of the algorithm outlined in this paper is illustrated using selected example events that span a range of different event types recorded on the UUSS network, including a moderate earthquake (a Mw 4.7 earthquake), a large open-pit mining explosion, a small mining-induced seismic event, and a regional event from outside the study region. By focusing on this limited dataset, we sacrifice generality with the goal of making the paper more illustrative of the breadth of issues that must be dealt with to use the WCEDS method to monitor a broad region. We verified the general applicability of our broad conclusions through analysis of many additional events recorded on the UUSS network.

2. An implementation of WCEDS as a regional event detector

This section describes an updated algorithm for implementing Equation 2 to detect events on regional seismic networks. The algorithm is based on our experience of adapting WCEDS to detect local and regional seismic events recorded on the UUSS seismic network. A complete listing of all the associated parameters that must be set is provided in Table 1, with the associated values for processing UUSS seismic data provided.

2.1 Overall framework

The overall processing flow in the updated WCEDS algorithm is illustrated in Figure 1. The following narrative discusses this figure by referencing the numbered labels, (1) - (5) in Figure 1.

Step 1: Initial Back-projection

Data from a regional network of sensors is processed in a moving back-projection window of duration t_{bp}; the start of the back-projection window corresponds to the u'th origin time hypothesis, τ_u (1). In each back-projection window, Equation 2 is applied to the network data in that window, $\underline{D}$, given some image matrix, $\underline{M}$, to produce the correlation matrix $\underline{X}$. Given a geographic distribution of event hypotheses in latitude and longitude – illustrated as a regular grid in (2), but not regular in practice, as described below – the power at each location, P, given the origin time corresponding to the start of the back-projection window, τ_u, is calculated. N_Δ (Equation 2) is calculated on basis of this distance threshold, and varies for each node in the network. For each origin time hypothesis, WCEDS stores the maximum power over all locations, $\max_{\lambda,\phi}\left(P_{\tau_u}\right)$.

Step 2: Peak Detection

As illustrated in Figure 1, event detection is applied in a time window, t_d, that lags the back-projection window (3) to identify peaks of $\max_{\lambda,\phi}\left(P_{\tau_u}\right)$ that exceed some threshold power, P_{thres} (dashed line in 4). As illustrated by (4), the purpose of using a trailing detection window is to ensure that individual values of $\max_{\lambda,\phi}\left(P_{\tau_u}\right)$ exceeding the threshold are peaks, through comparison with adjacent values. This requires two thresholds to be set, the threshold power, P_{thres}, and the minimum delay-time between peaks, t_{thres} (Table 1). While the specification of P_{thres} ensures a minimum threshold for triggering the detector, t_{thres} provides a contextual threshold that ensures that each peak is only detected once (this is equivalent to the use of a detrigger window, which is commonly used in seismic detectors). The peaks are stored as event detections, with the corresponding location and origin time written to a database.

Step 3: Phase elimination and Re-back-projection

The data are re-back-projected for each origin-time hypothesis (5) after removing the phases from any events that are detected from Steps 1 and 2. It is critical that all phases are removed to avoid false secondary events from forming; for Utah, even though we may only detect events using Pg, we remove Pg, Pn, and Lg phases for each event by removing data within -5 to +20 s of the predicted arrival time for each phase and filling the missing values by linear interpolation. The re-back-projection step repeats until no further events are detected given the thresholds used.

The purpose of processing the data using different distance thresholds is to ensure that both large network-wide events, and small sub-network-level events are both properly detected and located. Large events that are detected across a network need to be processed using a large distance threshold. If a small distance threshold is used for such events, or if distance weighting terms are used (as in Arrowsmith et al., 2016), it is possible to mislocate such events due to the inhomogeneous station distribution. False events can form on subnetworks by aligning the phases from large events at the incorrect event hypothesis. Conversely, small events, which may only be detected by a small number of stations, are not detected if using a large distance threshold, as their power is downweighted due to averaging in stations that have no observations because they are too far away (Equation 2). The phase elimination and re-back-projection step with a reduced distance threshold is a solution that enables us to detect and accurately locate both large and small events in the same study region.

2.2 Computation of travel times

To assess regional phase propagation in the study region, we stack STA/LTA transformed data using 8951 observations on vertical component seismograms from 77 $M_W > 2.5$ earthquakes in 2007 and 2008 (Figure 2). Clear Pg, Pn, and Lg arrivals are observed on the vertical components. The Pg and Pn travel times are calculated using the AK135 velocity model (Kennett, et al. 1995) and Tau-P toolkit (Crotwell, et al. 1999). Travel-time curves calculated using this model are superimposed on the stacked observations, indicating that the 1D Global model provides a reasonable fit to the observations, which are aligned based on event hypocentral locations

obtained by UUSS using a set of five 1D Earth models derived for different regions in Utah. The Lg travel time is calculated by assuming a constant group velocity of 3.5 km/s to fit the observed moveout of the Lg phase. The durations of the phase windows used in the stack, described below, are also illustrated in Figure 2. Because the Lg phase is relatively emergent, a longer phase window is used in the stacks, as described in more detail below.

2.3 Construction of grid of event hypotheses

The construction of a geographic grid of event hypotheses is necessary prior to implementation of WCEDS. In fact, the definition of the grid is critical for the successful implementation of WCEDS as it directly affects the detection thresholds and the false alarm rate. The relative effect of these two terms are discussed separately.

Detection thresholds

The capability of WCEDS to detect events requires the alignment of phases with an a-priori moveout for the 'correct' event hypothesis (i.e., the event hypothesis closest to the unknown true event location and time). For small events that are only detected by dense sub-networks, a dense set of grid nodes is required to ensure that the phases align sufficiently well to properly estimate the power of the event. If the grid nodes are too coarse, the event hypothesis closest to the truth may not result in adequate alignment of the phases. Without adequate alignment of phases, two consequences can occur: (1) we may not obtain a high enough estimate of power to detect the event at all, or (2) we may detect the event but obtain a poor location.

False alarm rate

False events can form from the phases of events that are outside the study area falsely correlating with the predicted phases for an event hypothesis grid point within the study area. Fortunately, such events only form from regional events outside the study area, and not for teleseismic events, as the arrival time delays associated with teleseismic arrivals cannot be fit to regional phase moveouts. There are three categories of possible approaches for preventing such events from contaminating our WCEDS event catalog.

1. Build the event outside the study area. The first approach is to build the long-range regional event outside the study area by including nodes outside the station network to absorb the phases from long-range events by building events that fit better there. This approach requires a modification to the image matrix, $\underline{M}$, to extend the range of times and distances far beyond the 200 km threshold used for initial back-projection (described in more detail below). Such a modification necessarily requires modeling multi-phases, and results in a significant increase in computation time through the increase in the dimension of the matrices in the dot-product computation (Equation 2).

2. Avoid building the event. A second approach is to avoid building false events. False events form where the phases of long-range regional events can be aligned sufficiently well with the predicted phases for an incorrect event hypothesis to trigger the detector. Such alignments are much less likely for grid nodes that have a good number of stations that are well-distributed azimuthally. Thus, eliminating grid nodes from locations that are poorly resolved by the network prevents WCEDS from building false events using phases from long-range regional events.

3. Build then reject the event. The third approach is to make various additional measurements on the waveforms to see if they are consistent with an event at the detected location to identify and eliminate false events after they have been built. Examples of screening criteria could include the use of polarization (Jurkevics 1988) or frequency information (Giudicepietro, et al. 2017).

We have found that avoiding building false events by removing grid nodes in areas of poor station coverage is the simplest solution to avoid false triggers on events from outside our study region. The downside of this solution is that events that occur in poorly-resolved regions will not be built.

To optimize the trade-off between algorithm speed and accuracy, we use a spatially-adaptive grid that is constructed as follows: (1) a high-resolution grid of nodes with a regular spacing in both latitude and longitude is defined over the geographic area covered by the network; (2) nodes with an azimuthal gap, defined as the maximum difference in the node-to-station

azimuths between azimuthally adjacent stations of <180° for stations within Δ_t km are discarded; (3) the node spacing is adjusted based on the distance to the N'th nearest station (a proxy for network density) using a function that relates distance to node spacing. By experimenting with different functions that relate the grid spacing to distance, we prefer a function of the form $\Delta g_{\min} + (\Delta g_{\min} + \Delta g_{\max})\left(1 - e^{-\beta d}\right)$, where d is the distance to the N'th nearest station, $(\Delta g_{\min}, \Delta g_{\max})$ are the minimum and maximum grid-spacings, and β is a constant that controls the rate of growth of the grid spacing with distance.

It is worth noting that, although we use a spatially-adaptive grid, we use a fixed time grid with spacing between successive origin-time hypotheses, δt, of 1 s. In principle, it would be preferable to update the temporal resolution based on the running estimate of power, allowing us to obtain finer resolution estimates of an event hypothesis without performing unnecessary additional computations when there is no evidence of an event. Building on this idea further, it may be desirable to adaptively update both the spatial and temporal resolutions of grid nodes on the basis of background power. These concepts are left for possible future research.

2.4 Construction of an image matrix

A second prerequisite to running WCEDS is the construction of an image matrix, represented in Equation 2 by $\underline{M}$. If constructing $\underline{M}$ using travel-time predictions through an Earth model (the 'synthetic implementation' of WCEDS), the matrix resolution must be specified by four parameters: $R_{\max}$, δR, $T_{\max}$, and δT (Table 1), which define the maximum distance and travel-time, as well as the discretization of distance and time in $\underline{M}$. These parameters need to be based on the maximum distance used in the back-projection step, Δ, and on the finest node resolution in the grid. One advantage of the synthetic implementation of WCEDS is that these parameters can be set at any value chosen by the user. The alternative implementation (the 'empirical implementation' of WCEDS) is to generate $\underline{M}$ using a set of a-priori events (this is the method used in Arrowsmith et al., 2016). The disadvantage of the empirical implementation is that these four parameters are limited by the data available.

3. Dataset and Pre-processing

This paper explores different implementations of WCEDS for event detection on the UUSS network in the state of Utah. The UUSS seismic network is comprised of a variety of different types of seismic instruments, including strong-motion, short-period, and broadband sensors. This study utilizes only the vertical components of the short-period and broadband stations (Figure 3). The UUSS network is designed for seismic hazard monitoring and subsequently follows the distribution of seismicity in the state, resulting in a spatially-variable network density. The grid node locations (Figure 3) were constructed following the method described above with the values of $\Delta_t, \Delta g_{\min}, \Delta g_{\max}, N$ defined in Table 1. The parameters were chosen by tuning the algorithm on example events to find a compromise between accuracy and computational speed.

This study focuses on four events, including a M_L 2.6 mine blast from the Bingham Canyon open-pit copper mine, a M_W 4.7 earthquake near Circleville, Utah, a small (magnitude not available) mining-induced seismic event associated with long-wall coal mining in central Utah (see Arabasz et al., (2005) for details on these events), and an event from outside the study region -- a M 3.5 event from the southern tip of the Salton Sea, California. These events span a range of sizes and occur both inside and outside the study region, and the events inside the study region occur in areas with different grid resolutions.

Prior to processing, we filter the vertical component data on all UUSS short-period and broadband seismometers using a bandpass Butterworth filter with passband from 0.5 – 4 Hz and 2 poles. We next process the data using an STA/LTA transform (Withers, et al. 1998) with STA window of 3 s and LTA window of 60 s (Table 1).

4. Application of synthetic WCEDS on Example Events

This section describes the application of WCEDS to the UUSS network data on our example events, using these events to explore the effect of different parameters. All the examples in this

section use a synthetically-generated image matrix, $\underline{M}$, that is based on travel-time predictions for the different regional phases: Pg, Pn, and Lg.

4.1 A Magnitude 4.7 earthquake

On 01/03/2011 12:06:36.00 UTC, a magnitude 4.7 earthquake occured at a depth of 5.4 km near Circleville, Utah (Arrowsmith, et al. 2012, Gammans 2013, Whidden and Pankow 2012). The earthquake was clearly recorded across the UUSS network, with evidence of Pg, Pn, and Lg phases on the vertical-component recordings (Figure 4).

Using this event, we explore the effect of different implementations of P phases in the image matrix on the WCEDS solution. We first apply WCEDS using the stack in Figure 5a, which models Pg out to 200 km. For a given distance in the image matrix, we construct a boxcar function that is zero before the predicted travel time, turning to one at the predicted travel time, and remaining at one for some finite duration – equivalent to M in Equation 1 – before returning to zero. The width of the boxcar function (δM in Table 1) is a parameter that must be tuned based on the data processing parameters used. For the stack in Figure 5a, we set δM = 4 s. The results of applying WCEDS to a one-minute period of data, centered approximately on the event origin time, are shown in Figure 6a. It is clear that there is a peak in the WCEDS power, $\max_{\lambda,\phi}\left(P_{\tau_u}\right)$, corresponding to the event, but that the peak is ~2s late compared with the estimated origin time published by UUSS. Further, a record section constructed for the WCEDS solution (Figure 6c) shows that the Pg arrivals are late compared with the predicted Pg time window given the WCEDS event hypothesis. We have found that WCEDS origin time solutions are often late unless we include pre-arrival downweighting terms in the image matrix. Figure 5b shows a stack with pre-arrival downweighting terms added via the inclusion of a time interval comprising -1's before the predicted travel time. This addition penalizes solutions where the actual onset time is early relative to the predicted onset time for a given event hypothesis (any power in this window is subtracted from the total estimate of power). The result obtained by applying WCEDS using the new stack in Figure 5b are shown in Figures 6b and 6c. The location solution obtained with downweighting terms included in the image matrix more closely matches to the

origin time reported by UUSS (Figure 6b), which appears to fit the observations better (Figure 6c). Furthermore, the precision of the estimate is better (i.e., the peak is narrower). Because the use of pre-arrival downweighting results in consistent improvements in event hypothesis estimates, we use it in all subsequent analysis in this paper.

The third implementation includes Pn, and extends the maximum distance in the image matrix ($R_{\max}$ in Table 1) to 450 km (Figure 5c), with Δ also set to 450 km in order to capture arrivals across the whole network. In this image matrix, the first arrival is either Pg or Pn, with pre-arrival downweighting terms added. Beyond ~200 km, the Pn and Pg arrivals separate into distinct windows. As shown in Figure 7, the best solution obtained with this third implementation is identical to the best solution obtained using Pg alone (with downweighting terms added) inside 200 km. However, the solution is less precise (the peak is broader in both time and space) and the computation time is increased in this third implementation. This result suggests that using the whole network does not provide any improvement for detecting large events over using a distance threshold, Δ, of 200 km. On the basis of this, for subsequent analyses, we set Δ=200 km (and Δ=75 km, as discussed in more detail below). By setting the larger distance threshold of Δ=200 km, the first arriving energy is Pg (i.e., the distance is within the Pg/Pn crossover distance). Within 200 km, Pg consists of diving waves in the crust, making accurate computation of the travel time, given a 1D or 3D model, relatively straightforward using ray theory (Kennett 2002). In addition to the fact it doesn't appear to improve the results, increasing the distance threshold requires us to account for the Pg/Pn crossover, and for complex multi-pathing effects, which are not adequately captured with ray theory.

4.2 Open-pit mine blast

The second event is an open-pit mine blast from the Bingham Canyon open-pit copper mine that is listed in the USGS Comcat with origin time 2014/11/23 23:35:10, location 40.537°N, -112.1384°E, and magnitude, M_L=2.6. The event was confirmed as a blast by UUSS analysts and reported to the USGS because it exceeded a magnitude of 2.5. A record section of the event (Figure 8) shows observable arrivals out to ~350 km with weaker Lg arrivals than were observed

for the Circleville earthquake (Figure 4). We note that the stack of STA/LTA transforms for 77 earthquakes with M_L >2.5 using the same STA/LTA parameters shows clear Lg arrivals (Figure 2), indicating that Lg is common for other types of event in this region. Using the regional stack to define a suitable Lg window (defined using a group velocity of 3.5 km/s and with δM=8 s), we explore the effect of adding Lg to the image matrix for this event.

The results, shown in Figure 9, illustrate that the addition of Lg phases, while not affecting the peak estimate of the event location and origin time, has the effect of reducing the precision of the event hypothesis estimate. This result suggests that adding Lg to the stack, when Lg is not present in the data, reduces the resolution of WCEDS as an event detector. Because small mining-related events can generate very little shear-wave energy, we prefer a simple Pg implementation to the inclusion of both Pg and Lg in the stack for general monitoring purposes, i.e. for detecting both earthquakes and mining explosions. Finally, we note that the WCEDS solution, using either stack in this case, leads to an event hypothesis of 40.53°N, -112.15°E, and 2014/11/23 23:35:11 UTC. The solution is within 1.25 km of the UUSS event location, and is also located inside the Bingham Canyon pit.

4.3 A small mining-Induced seismic event

The third event is a mining-induced seismic (MIS) event from the Book Cliffs area (Arabasz, et al. 2005) that is not listed in the UUSS catalog. Due to a high level of mining-related seismicity in the Book Cliffs area, UUSS use a high event-detection threshold in that region. Applying WCEDS to data from the UUSS network detects many events in this region that are not listed in the UUSS event catalog (Arrowsmith et al., 2016). The event discussed here was detected by WCEDS with Δ set to 75 km with an origin time of 2011/01/03 09:10:40 and location of 39.62°N, 110.40°W. Given the WCEDS event hypothesis, a record section is shown in Figure 10.

This event provides a useful illustration of the importance of specifying a suitable grid before running WCEDS. The event is located in a dense subcluster of stations, and high resolution event hypothesis grid locations are needed to ensure proper alignment of the phases across the

nearby stations. This is illustrated in Figure 11, which shows a comparison between the WCEDS solution using the high-resolution grid in Figure 3, and the solution obtained using a relatively low-resolution grid, with fixed grid spacing of 0.1×0.1°, as was used by Arrowsmith et al. (2016). The figure shows that the MIS event solution is very different between the two runs. The event is correctly located within the dense subcluster of stations in the Book Cliffs area using the high-resolution node locations shown in Figure 3, but is incorrectly located to the east of the region when using the low-resolution set of grid nodes. In the latter solution, there are two problems: (1) the resolution of grid nodes inside the Book Cliffs region is insufficient, causing the estimates of power in that region to be less than when using a high-resolution set of nodes, and (2) the inclusion of grid nodes that are poorly resolved by the network (i.e., have a poor azimuthal gap) results in high estimates of power due to the fortuitous alignment of high-amplitude phases with an event in that location at a different time from the true event.

4.4 A M 3.5 event from Southern California

The final event studied in this section is a M 3.5 earthquake from outside the study area, located in California, on the south shore of the Salton Sea, by the USGS with location of 33.16°N, -115.644°E, and origin time of 2011/01/01 00:34:10. If using an overly-dense set of poorly-resolved grid nodes, as shown in Figure 12, it is possible to falsely detect and locate this event inside the study region by fortuitous alignment of the phases with the incorrect event hypothesis. However, by using a less dense set of only well-resolved grid nodes (Figure 3), the event is not detected and barely results in an increase in the background power. As shown in Figure 12, false solutions with high power corresponding to the far regional event occur in poorly resolved regions of the network, in the SW and SE corners of the study area.

5. Comparing synthetic and empirical implementations

As discussed in the Introduction, WCEDS can be implemented by either (a) synthetically generating the image matrix, as is explored in the previous section, or (b) by generating the image matrix by stacking observed data (e.g., Arrowsmith et al., 2016). The latter approach assumes that one has access to a prior earthquake catalog and corresponding waveform data,

enabling one to stack STA/LTA transformed data from multiple events in time versus distance bins. The parameters used for implementing WCEDS in this paper (Table 1) and associated grid node locations (Figure 3) require a higher-resolution empirical stack than was developed in Arrowsmith et al. (2016) using ~9000 observations. In fact, the ability to quickly and easily configure the image resolution is a clear advantage of the synthetic implementation explored in this paper. When developing an image matrix from empirical data, the image resolution possible is directly controlled by the number and distribution of source-receiver paths. As a consequence of this, we do not compare the solutions above with equivalent solutions obtained using an empirical implementation of WCEDS. However, to show that the two approaches provide equivalent results when run at the same resolution, we compare a 1-hour time period processed using both empirical and synthetic image matrices (Figure 13). The 1-hour time period included the mainshock of the Circleville earthquake (see Section 4.1), and the initial sequence of aftershocks. The image parameters are as in Table 1 but with the resolution on distance, δR, set to 5 km instead of 0.1 km in order to provide adequate averaging of observations in the empirical image matrix. A consequence of this lower-resolution image matrix is that the fine grid resolution shown in Figure 3 does not provide any advantages over a coarser resolution grid. Consequently, the grid is set at the $0.1 \times 0.1^{\circ}$ grid used in Arrowsmith et al. (2016). Because the empirical stack necessarily includes all phases, Pn, Pg, and Lg phases are included in the synthetic stack without pre-phase downweighting terms added (these are not present in the empirical stack) to make the comparison as equal as possible. While, as discussed above, improved solutions are possible using the synthetic implementation with different parameters, we wish to show that both implementations provide equivalent results with the same choice of parameters. The results, shown in Figure 13, confirm this hypothesis. Despite the fact that the synthetic stack does not include the amplitude information that is present in the empirical stack, the results are essentially equivalent in terms of the event detections that are obtained with comparable thresholds. However, while this shows that it is possible to get comparable results using empirical or synthetic implementations, most of the improvements discussed above cannot be realized with the empirical implementation. Thus, to perform high-resolution event detection with WCEDS, we recommend using the synthetic implementation.

7. Conclusions

Building catalogs of events automatically is a fundamental task in routine earthquake or explosion monitoring applications. For most local and regional applications, the standard approach for tackling this problem involves separate station-level (detection, onset time estimation) and network-level (association, phase estimation, and location) processing tasks. This classical 'pick-based' approach has become the conventional technique used at most monitoring centers as it works relatively well for sparse networks, and requires relatively little computational power. However, as network densities increase, and computational resources become less constrained, backprojection techniques like WCEDS offer some compelling advantages. The first advantage is simplicity; without separating station and network level tasks, multiple tasks are accomplished in fewer steps. Second, backprojection techniques may require fewer analyst resources (e.g., onset-time picks or phase labels don't need to be adjusted manually when reviewing events), since the concept of a pick is not necessary. However, as is shown in this paper, there are a number of practical considerations involved in implementing a backprojection approach like WCEDS operationally.

This paper has explored a general backprojection approach, WCEDS, that can be applied in two distinct ways: (1) synthetically (an approach that is equivalent to reverse-time migration using a 1D model), and (2) empirically (e.g., Arrowsmith et al., 2016). While both approaches are shown to produce similar results when they are tuned to be equivalent, this paper has outlined a number of strategies for obtaining high-resolution results using the synthetic approach, which are not possible using the empirical approach. To implement high-resolution WCEDS for event detection, we have developed an algorithm for implementing the basic WCEDS equation (Equation 2) on heteorogeneous seismic networks that can detect events of different types, sizes, and locations relative to the station constellation. We have shown that an iterative approach that uses different distance thresholds is necessary for robust detection of both large and small events, and that a spatially-variable grid with only well-resolved nodes prevents artifacts associated with events outside the region from forming. For detecting events on the

UUSS network, our findings lead to the conclusion that a synthetic implementation of WCEDS, using only the Pg phase out to 200km, but including pre-arrival downweighting terms, is our preferred approach.

Acknowledgements

Sandia National Laboratories is a multi-program laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidary of Honeywell International LLC, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

Arabasz, W. J., S. Nava, M. K. McCarter, K. L. Pankow, J. C. Pechmann, J. Ake, and A. McGarr (2005). Coal-Mining Seismicity and Ground-Shaking Hazard: A Case Study in the Trail Mountain Area, Emery County, Utah, Bull. Seism. Soc. Am. **95** 18-30.

Arrowsmith, S., R. Burlacu, K. L. Pankow, B. Stump, R. Stead, R. W. Whitaker, and C. Hayward (2012). A Seismoacoustic study of the 2011 January 3 Circleville earthquake, Geophys. J. Int.

Arrowsmith, S., C. Young, S. Ballard, M. Slinkard, and K. L. Pankow (2016). Pickless Event Detection and Location: The Waveform Correlation Event-Detection System (WCEDS) Revisited, Bull. Seism. Soc. Am. **5** 2037 - 2044.

Baker, T., R. Granat, and R. W. Clayton (2005). Real-time Earthquake Location Using Kirchoff Reconstruction, Bull. Seism. Soc. Am. **95** 699 - 707.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP Toolkit: Flexible Seismic Travel-time and Ray-path Utilities, Seism. Res. Lett. **70**.

Gammans, C. N. L. (2013). Low-angle normal faulting in the Basin and Range-Colorado Plateau transition zone during the January 3, 2011 Circleville, UT earthquake sequence, University of Utah, 69.

Giudicepietro, F., A. M. Esposito, and P. Ricciolino (2017). Fast Discrimination of Local Earthquake Using a Neural Approach, Seism. Res. Lett. **88** 1989 - 1096.

Jurkevics, A. (1988). Polarization analysis of three-component array data, Bull. Seism. Soc. Am. **78** 1725 - 1743.

Kao, H., and S.-J. Shan (2004). The Source-Scanning Algorithm: mapping the distribution of seismic sources in space and time, Geophys. J. Int. **157** 589 - 594.

Kennett, B. (2002). The Seismic Wavefield Volume II: Interpretation of Seismograms on Regional and Global Scales, Cambridge University Press.

Kennett, B., E. Engdahl, and R. Buland (1995). Constraints on seismic velocities in the Earth from travel times, Geophys. J. Int. **122** 108 - 124.

Le Bras, R., and J. Wuster (2002). IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data, in *IDC Documentation*.

Shearer, P. M. (1994). Global seismic event detection using a matched filter on long-period seismograms, J. Geophys. Res. **99** 13713 - 13725.

Whidden, K. M., and K. L. Pankow (2012). A Catalog of Regional Moment Tensors in Utah from 1998 to 2011, Seism. Res. Lett. **83**.

Withers, M., R. Aster, and C. Young (1999). An Automated Local and Regional Seismic Event Detection and Location System Using Waveform Correlation, Bull. Seism. Soc. Am. **89** 657 - 669.

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. Seism. Soc. Am. **88** 95 - 106.

Young, C., J. Beiriger, S. Moore, J. Trujillo, and M. Harris (1998). The Waveform Correlation Event Detection System Project, Phase II: Testing with the IDC Primary Network, Sandia Report.

Young, C., J. Beiriger, J. Trujillo, M. Withers, R. Aster, L. Astiz, and P. M. Shearer (1996). WCEDS: A Waveform Correlation Event Detection System, in *DOE Report*.

Young, C., M. Harris, J. Beiriger, S. Moore, J. Trujillo, M. Withers, and R. Aster (1996). The waveform correlation event detection system project phase 1: issues in prototype development and testing, Sandia National Laboratories.

Tables

Parameter type	Parameter	Description	Value
Filter parameters	$f_{\min}$	Low-frequency cutoff of the Butterworth filter (Hz)	0.5
	$f_{\max}$	High-frequency cutoff of the Butterworth filter (Hz)	4.
	STA	STA duration (s)	3.
	LTA	LTA duration (s)	60.
Image parameters	$R_{\max}$	Maximum distance in the image matrix (km)	200
	δR	Resolution of distance in the image matrix (km)	0.1
	T_{max}	Maximum travel-time in the image matrix (s)	100
	δT	Resolution of travel-time in the image matrix (s)	0.05
	δM	Half width of time window around predicted arrival (s)	*2.*
Event hypothesis grid parameters	δt	Time increment between successive origin time hypotheses (s)	1.
	Δ_t	Threshold distance for estimating azimuthal gap (kilometers)	75
	N	N'th furthest station used for estimating node spacing	6
	$\Delta g_{\min}$	Minimum spacing between grid nodes (°)	0.005
	$\Delta g_{\max}$	Maximum spacing between grid nodes (°)	0.1
	Δ	Maximum node-station distance for calculating N_Δ in Equation 2 (km)	200 or 75
Event detection parameters	P_{thres}	Threshold power for event detection	0.3
	t_{thres}	Threshold time between peaks (s)	60.
	t_d	Duration of detection window (s)	120.

Table 1. A complete listing of all parameters that must be set for applying WCEDS to detect events, with the corresponding parameters used for the UUSS data analysis provided. Note that δM is only required if running WCEDS using a synthetic image matrix.

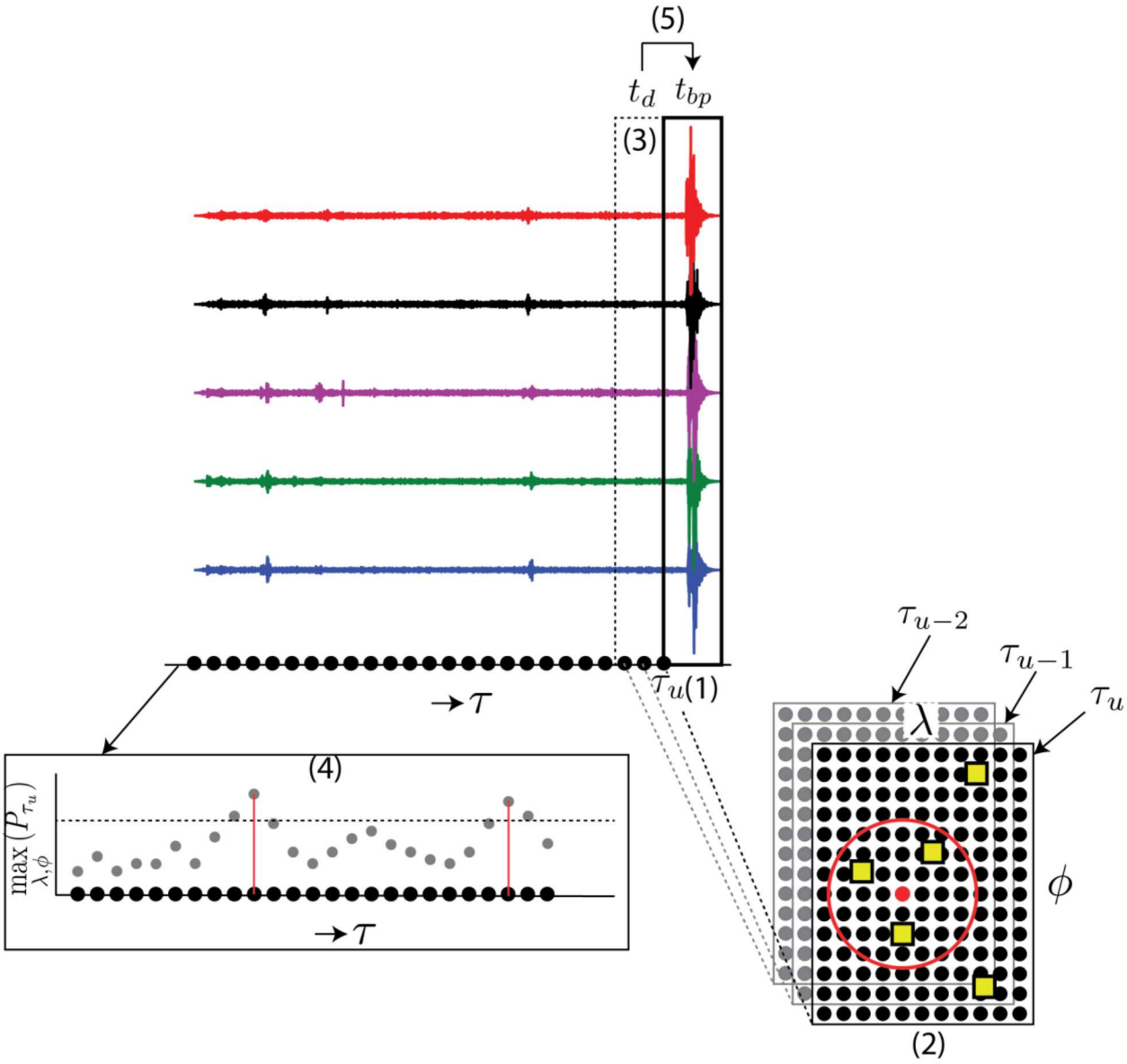

Figure 1. Cartoon illustrating the implementation of WCEDS as an event detector for a hypothetical five-station network. Data from the five stations flow into the WCEDS pipeline from left to right. WCEDS works by back-projecting data over a range of origin times, τ, where for the *u*'th origin time, τ_u, WCEDS back-projects the data in a window of duration t_{bp} (1) for a set of discrete spatial locations, (ϕ, λ) (2). For each origin time, WCEDS evaluates the peak power at each spatial location and retains the maximum power and corresponding spatial location of maximum power. By searching for peaks in a trailing detection window (3), WCEDS finds peaks above a user defined detection threshold (4). Each peak constitutes a detected event, with the corresponding origin time and spatial location written to an event catalog. After detecting events, a final feedback loop removes the phases from those events from the data and repeats the back-projection step (5) until no further events are found above the threshold.

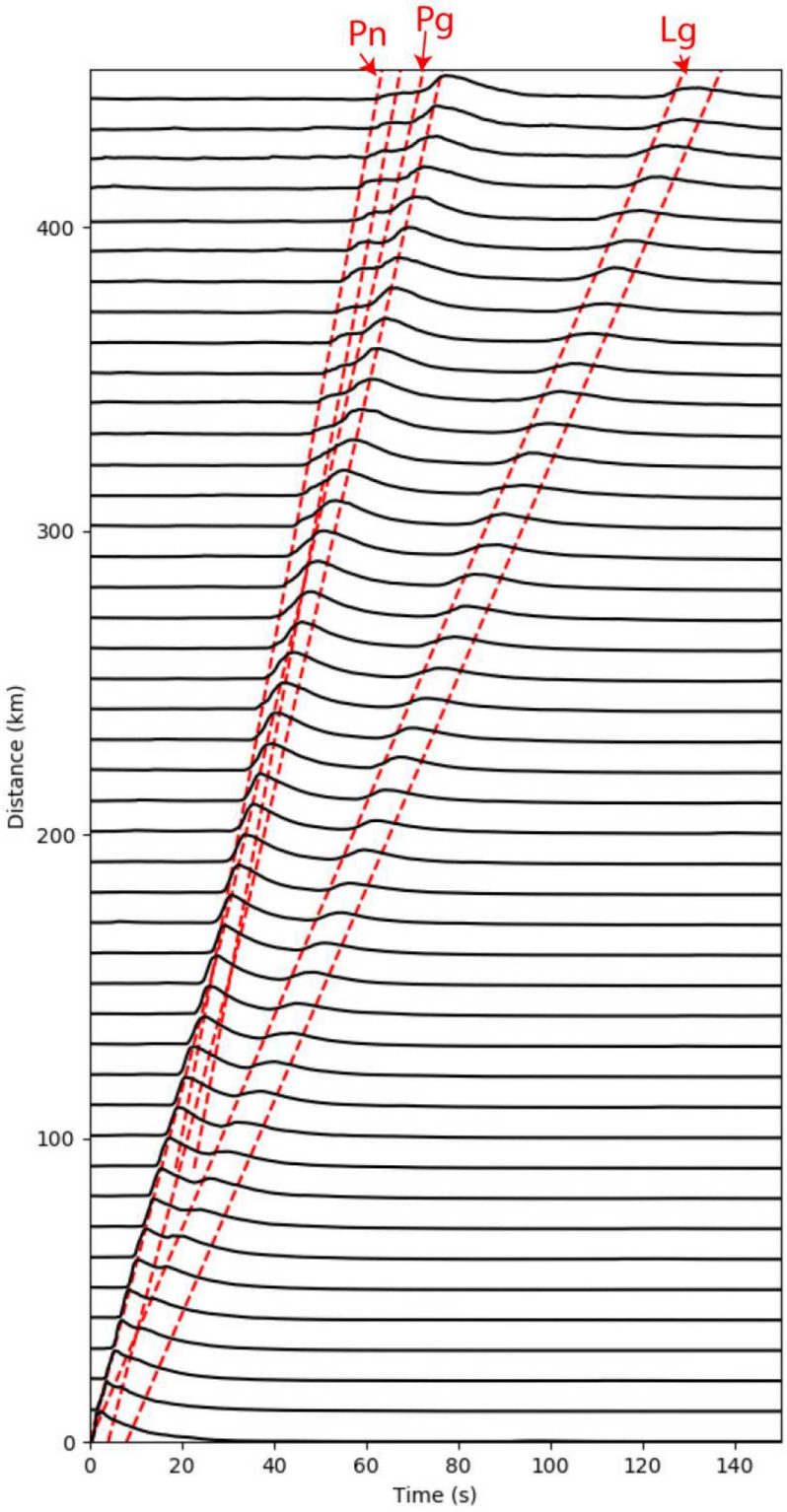

Figure 2. Record section created from a stack of 8951 observations, transformed using the STA/LTA filter used in this paper, from 77 M_W > 2.5 events occuring in Utah in 2007 and 2008. Earthquake origin times and locations are obtained from the UUSS seismic catalog. Dashed red lines show the predicted arrival times plus δM for Pg, Pn, and Lg that are used to generate the synthetic stacks in this study (based on the AK135 model).

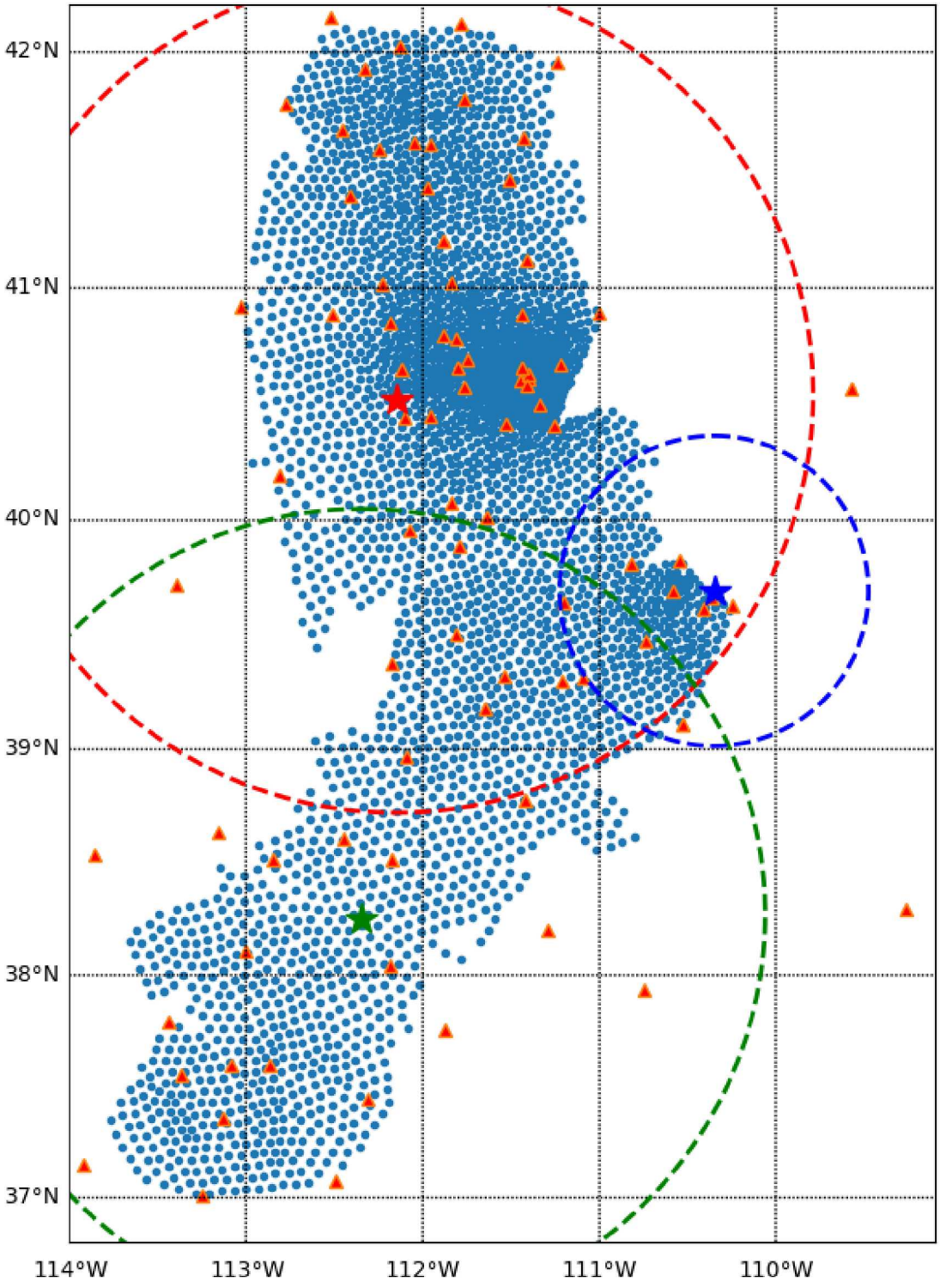

Figure 3. Locations of seismic stations (orange triangles) and grid nodes (blue dots) used in this study. The locations of three of the events explored in this study, the Bingham copper mine blast, the Circleville earthquake, and the mining-induced seismic event are shown by red, green, and blue stars; dashed circles denote locations at Δ_t km from each event (with radius of 200 km for both the Circleville earthquake and Bingham mine events, and 75 km for the MIS event).

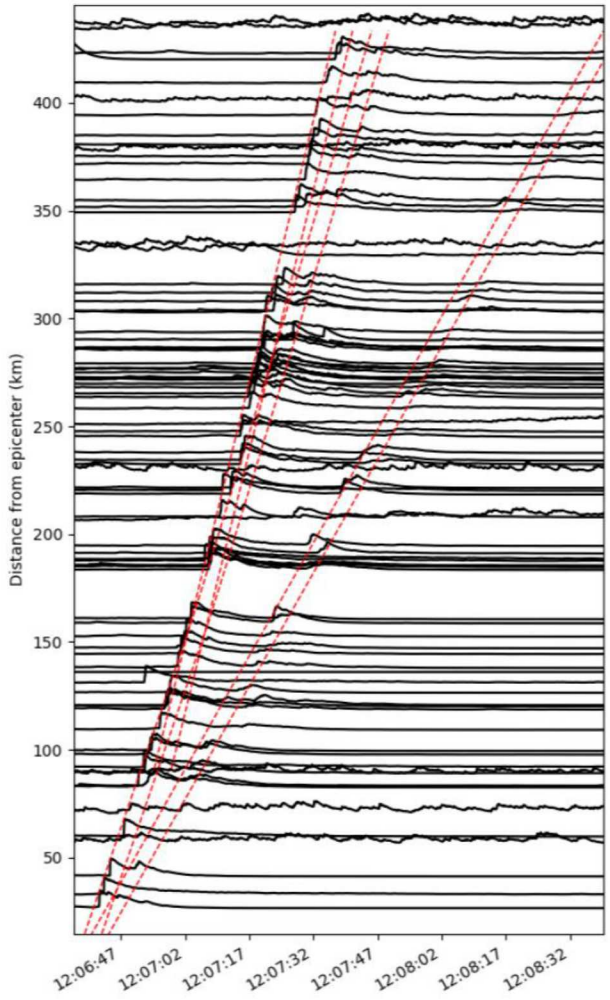

Figure 4. Record section for the M 4.7 Circleville earthquake showing STA/LTA transformed filtered data. Red dashed lines show the predicted arrival time windows of Pg, Pn, and Lg using the AK135 velocity model with δM=2.

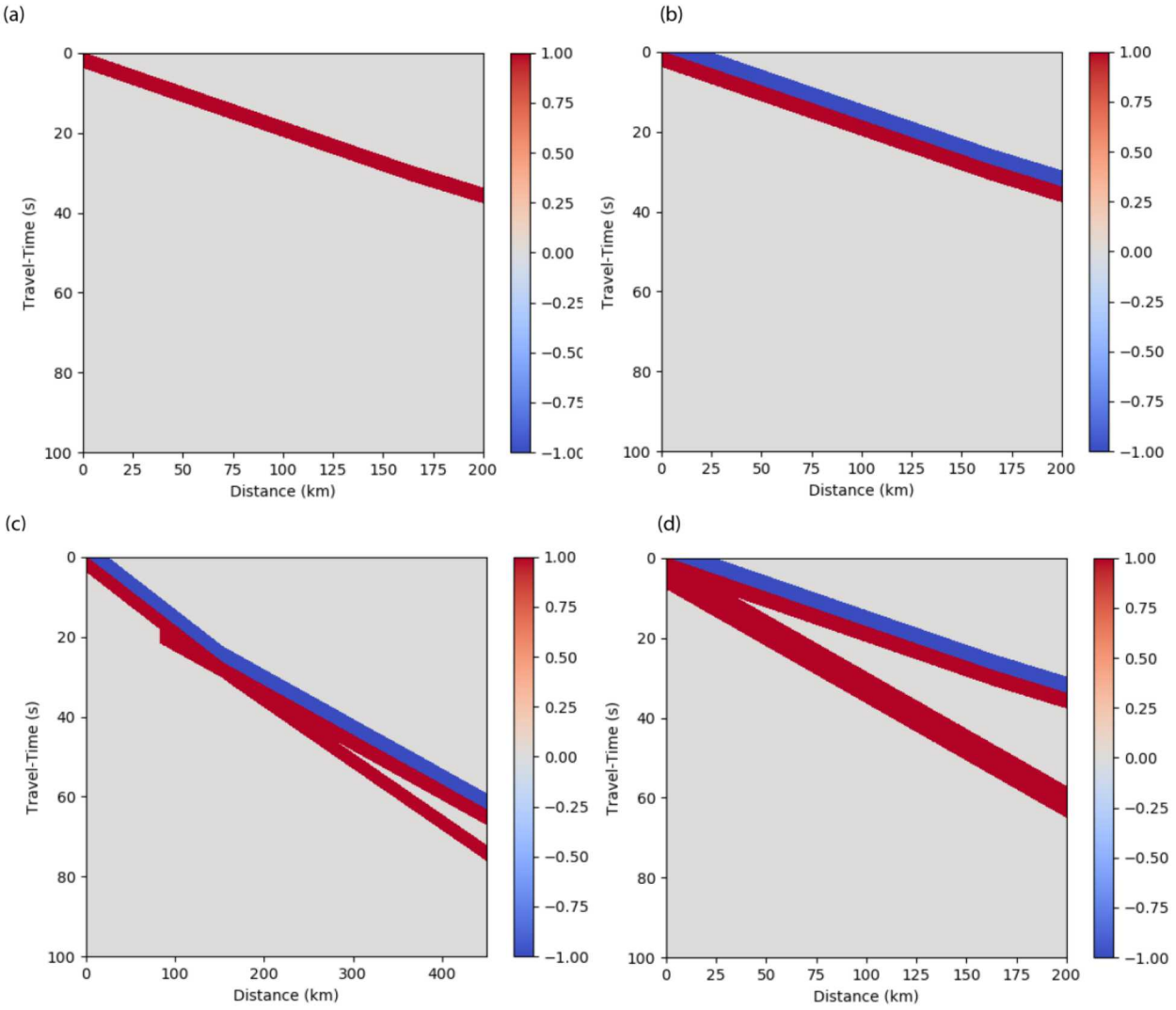

Figure 5. Image matrices used for computing the WCEDS solutions in subsequent figures. From left-to-right, the matrices are: (a) Pg predictions only, (b) Pg predictions with pre-arrival -1's to penalize late onset times, (c) Pg+Pn predictions with pre-arrival -1's out to 450 km, (d) Pg+Lg predictions with pre-arrival -1's out to 200 km.

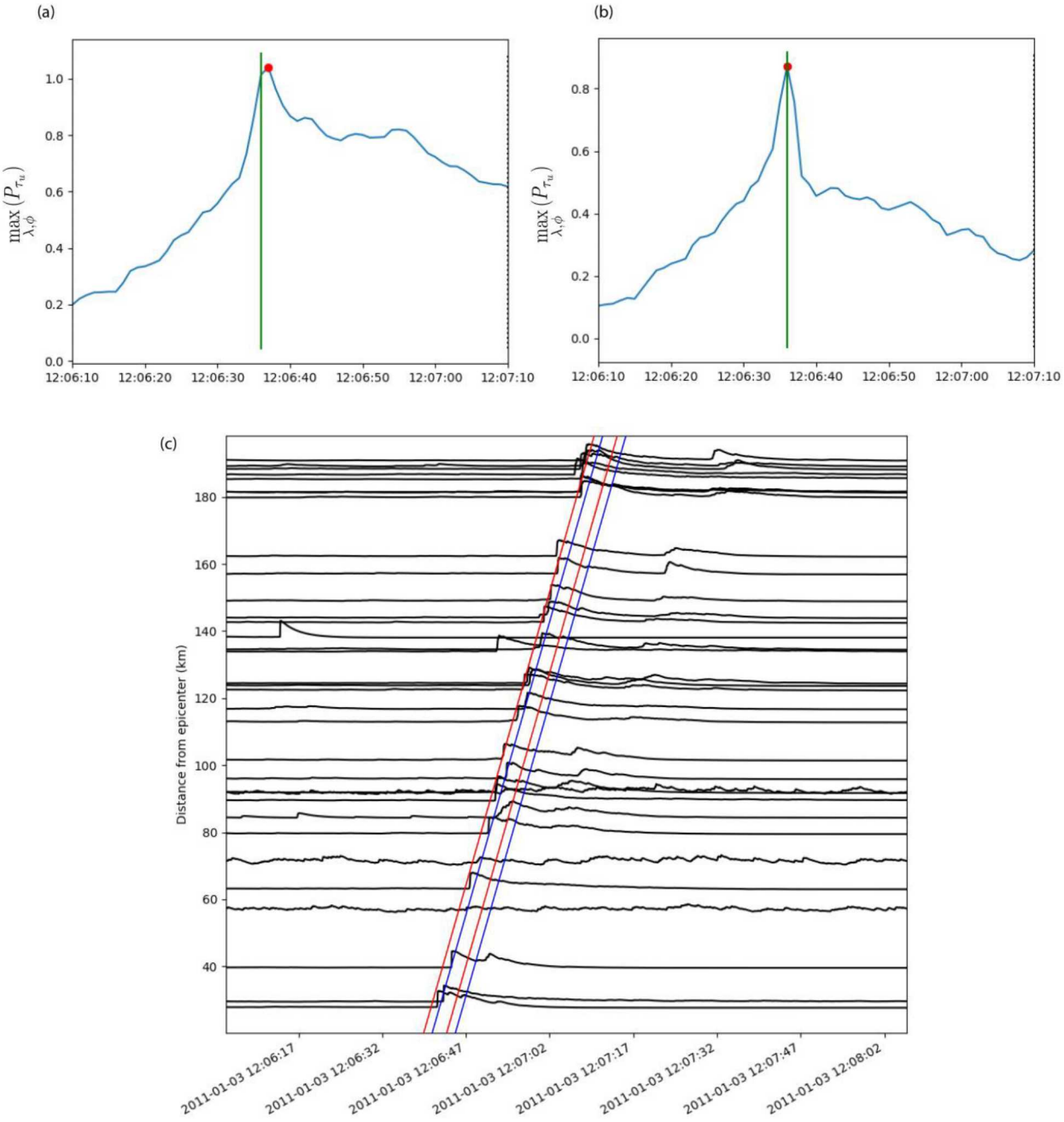

Figure 6. Comparison between the WCEDS origin-time solutions obtained with the stacks shown in Figures 5a and Figure 5b and all other parameters the same (Table 1). (Top) The blue line shows the temporal variation of peak power, $\max_{\lambda,\phi}(P_{\tau_u})$, obtained for a 1-minute time interval using the stack in Figure 5a (left) and the stack in Figure 5b (right). The red dot denotes the WCEDS solution, corresponding to the peak power. The vertical green line denotes the UUSS estimate of the origin time. (Bottom) Record section showing the STA/LTA processed data with predicted arrival times and predicted arrival times + δM for the solution obtained for the stack in Figure 5a (blue lines) and the solution obtained for the stack in Figure 5b (red lines).

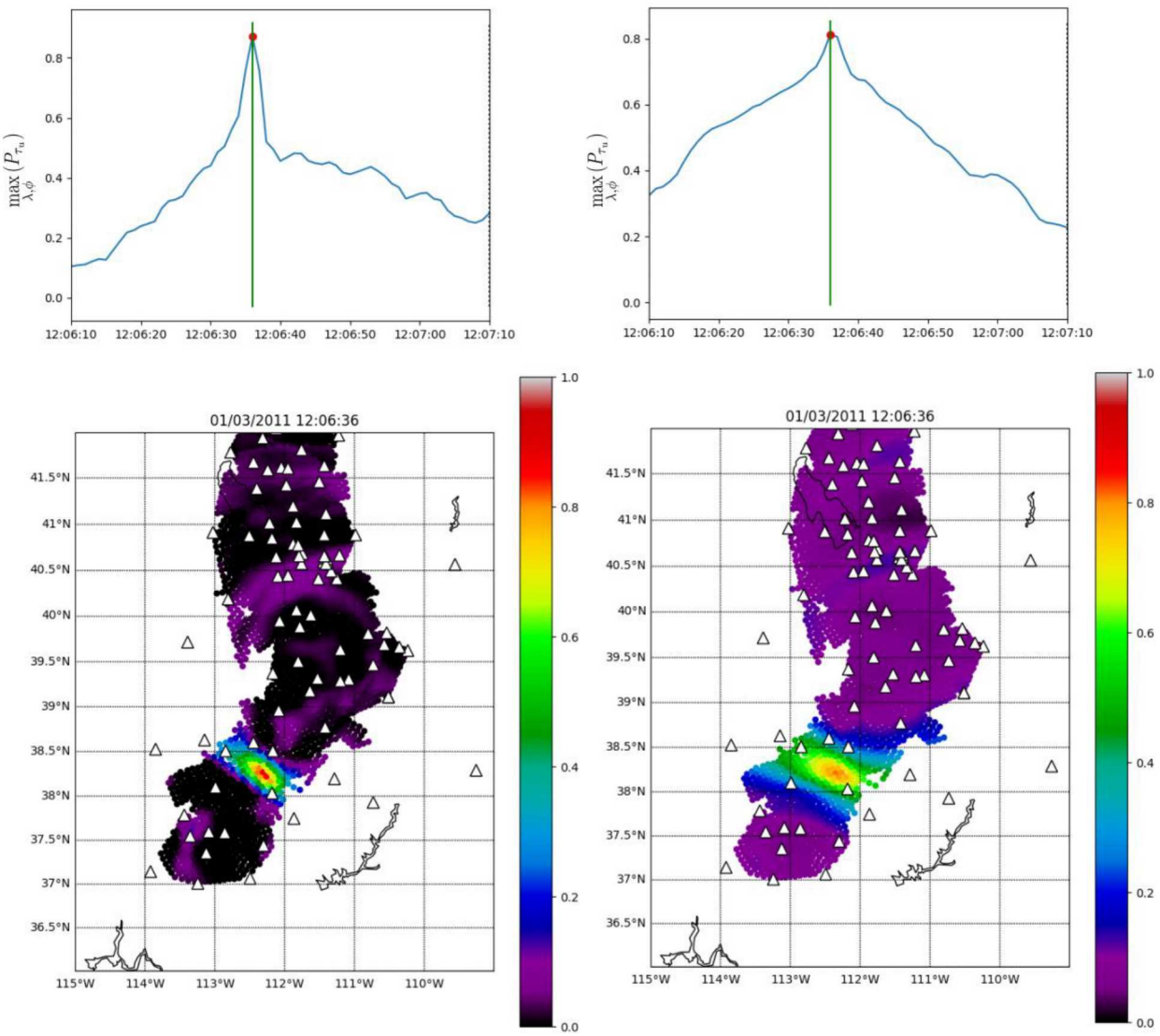

Figure 7. Comparison between the resolution in time and space using the stack in Figure 5b (left panel) with the stack in Figure 5c (right panel). (Top) The blue line shows the temporal variation of peak power, $\max_{\lambda,\phi}(P_{\tau_u})$, obtained for a 1-minute time interval. (Bottom) The spatial variation in power (color scale) at the time corresponding to maximum power.

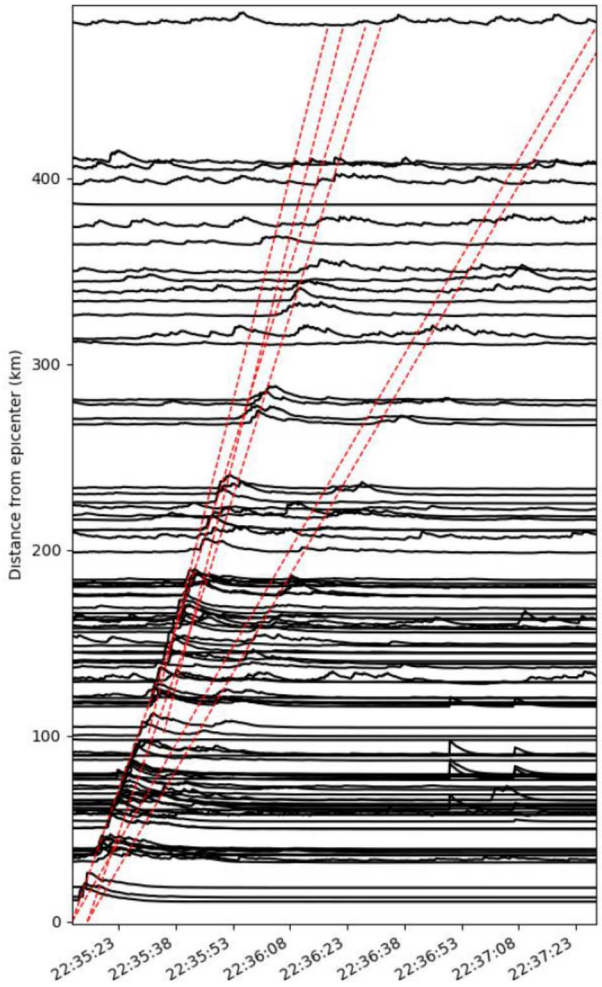

Figure 8. Record section for the M 2.6 Bingham Canyon mine blast showing and STA/LTA transformed filtered data. Red dashed lines show the predicted arrival time windows of Pg, Pn, and Lg using the AK135 velocity model with δM=2.

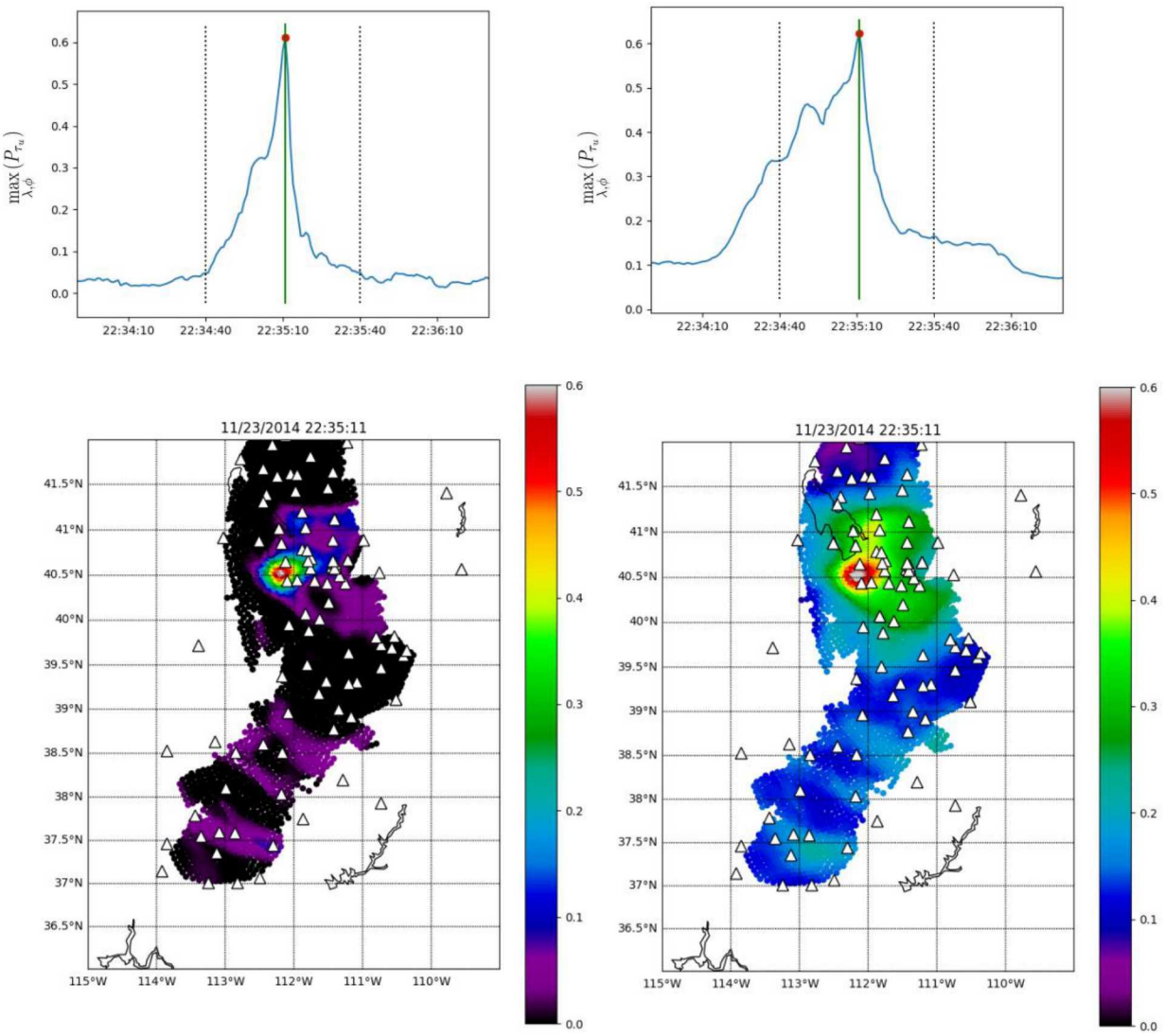

Figure 9. Comparison between the resolution in time and space using the stack in Figure 5b (left panel) with the stack in Figure 5d (right panel). (Top) The blue line shows the temporal variation of peak power, $\max_{\lambda,\phi}(P_{\tau_u})$, obtained for a 1-minute time interval. (Bottom) The spatial variation in power (color scale) at the time corresponding to maximum power.

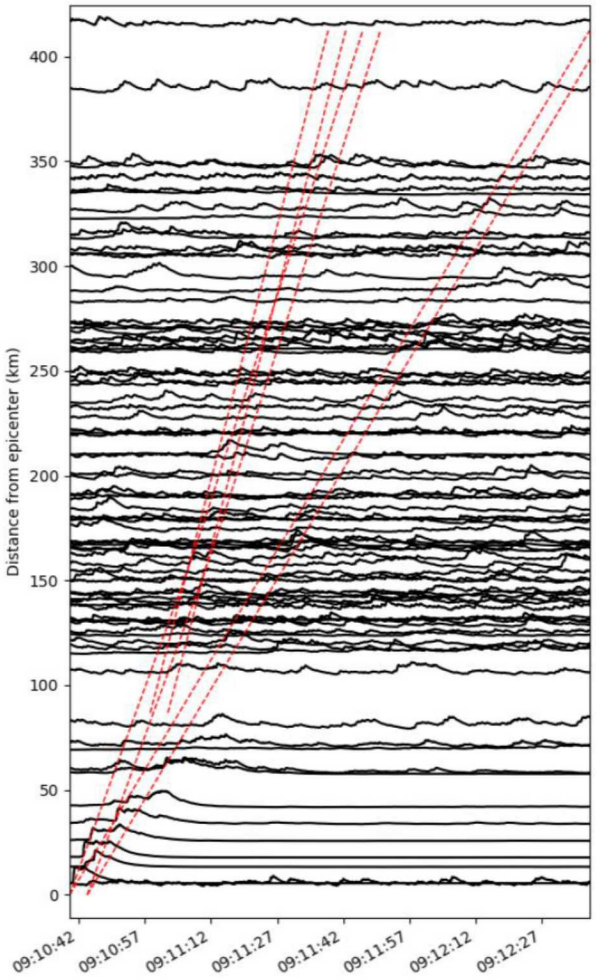

Figure 10. Record section for the mining-induced seismic event showing STA/LTA transformed filtered data. Red dashed lines show the predicted arrival time windows of Pg, Pn, and Lg using the AK135 velocity model with δM=2.

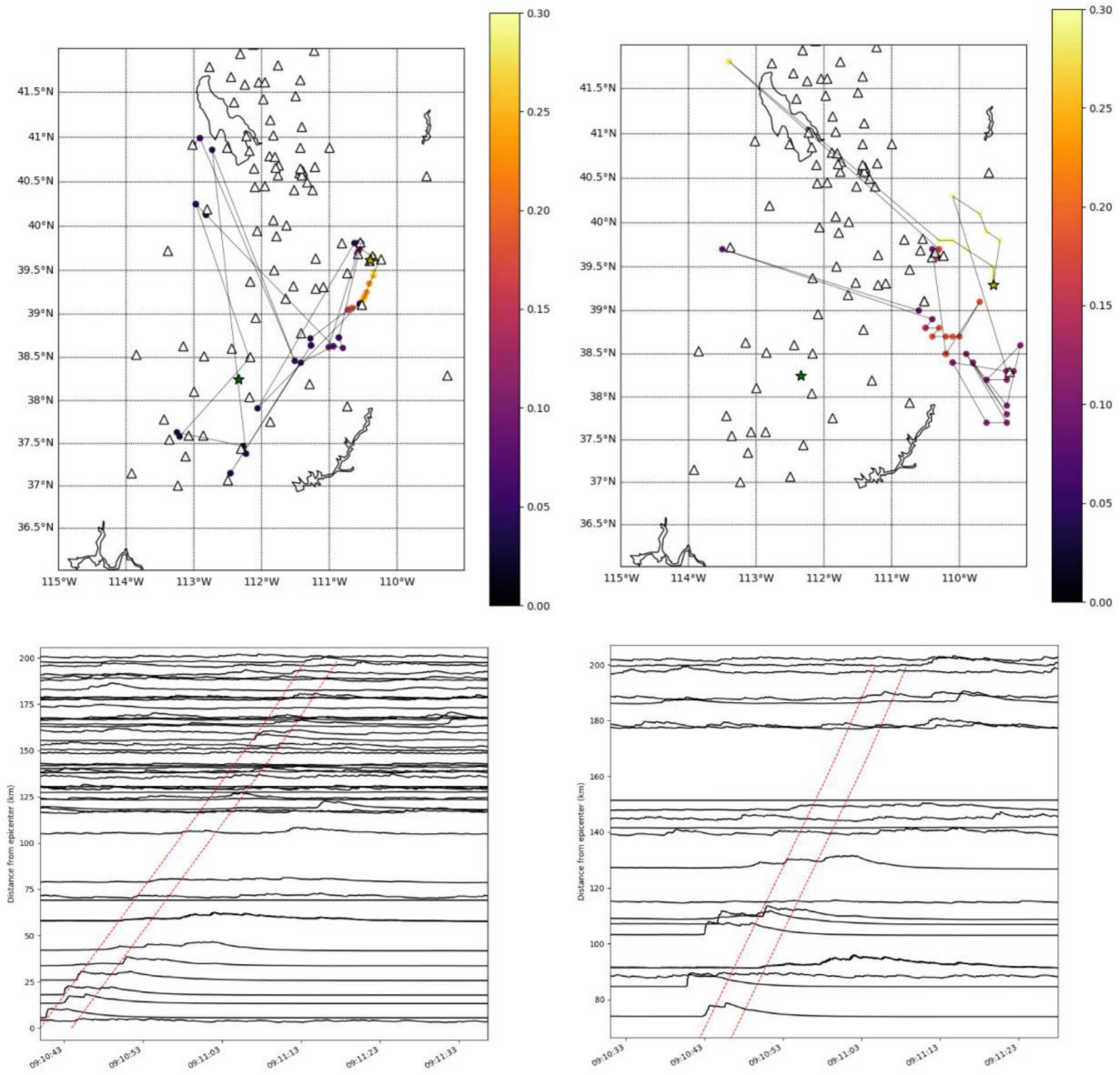

Figure 11. WCEDS solutions for the mining-induced seismic event using (left) the node locations shown in Figure 3, and (right) a low-resolution set of node locations with fixed spacings between nodes of 0.1° in both latitude and longitude. The top panel shows the locations of sequential estimates of $\max_{\lambda,\phi}(P_{\tau_u})$ in a 1-minute time period centered on the event origin time estimate (color-coded by power), and the bottom panel shows record sections for the event hypothesis corresponding to the peak WCEDS power in the time interval (shown by the stars in the maps on the top panel).

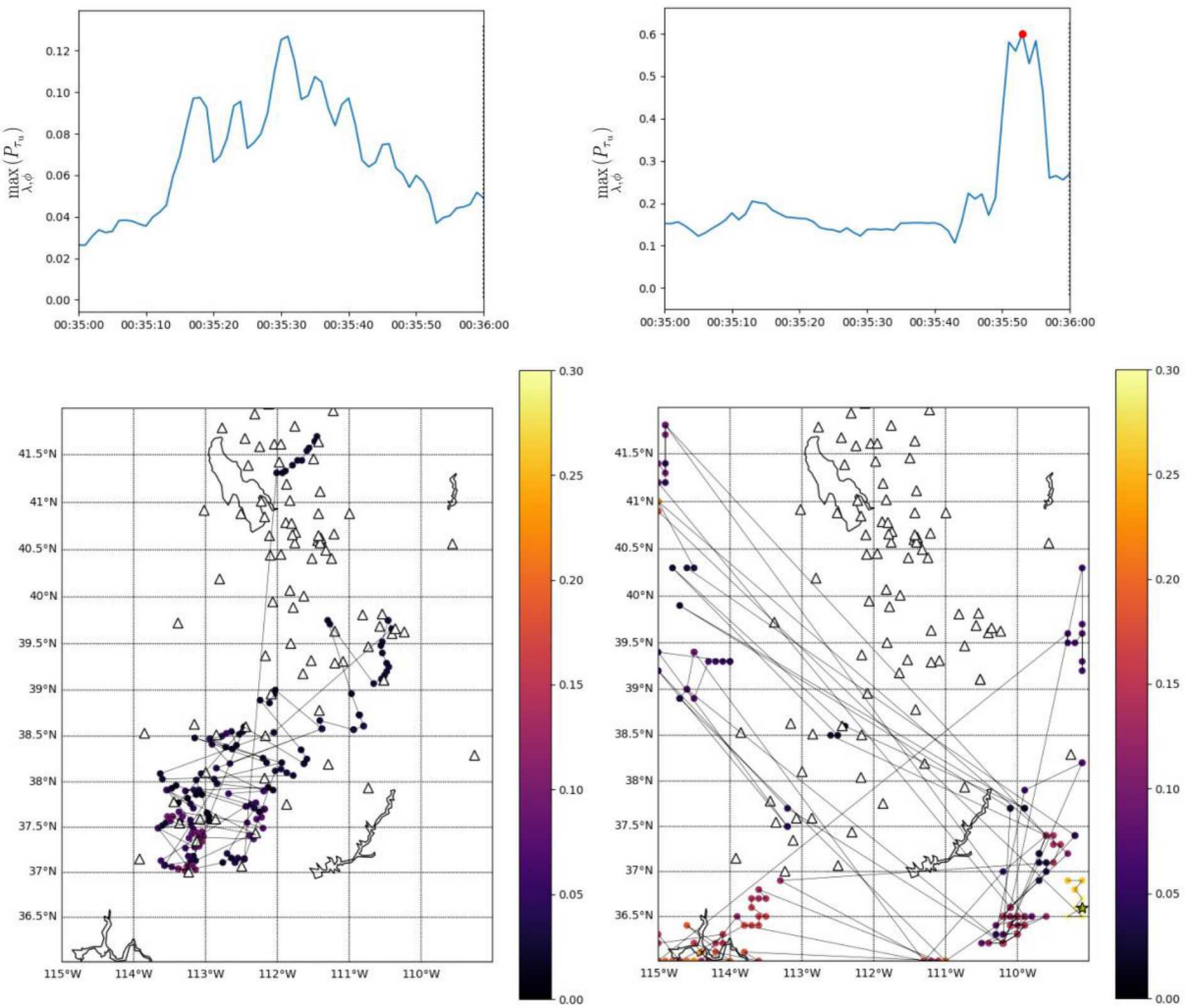

Figure 12. WCEDS solutions for the M 3.5 California event using (left) the node locations shown in Figure 3, and (right) a low-resolution set of node locations with fixed spacings between nodes of 0.1° in both latitude and longitude. (Top) The blue line shows the temporal variation of peak power, $\max_{\lambda,\phi}(P_{\tau_u})$, obtained for a 1-minute time interval. (Bottom) the locations of sequential estimates of $\max_{\lambda,\phi}(P_{\tau_u})$ in a the same 1-minute time inveral (color-coded by power).

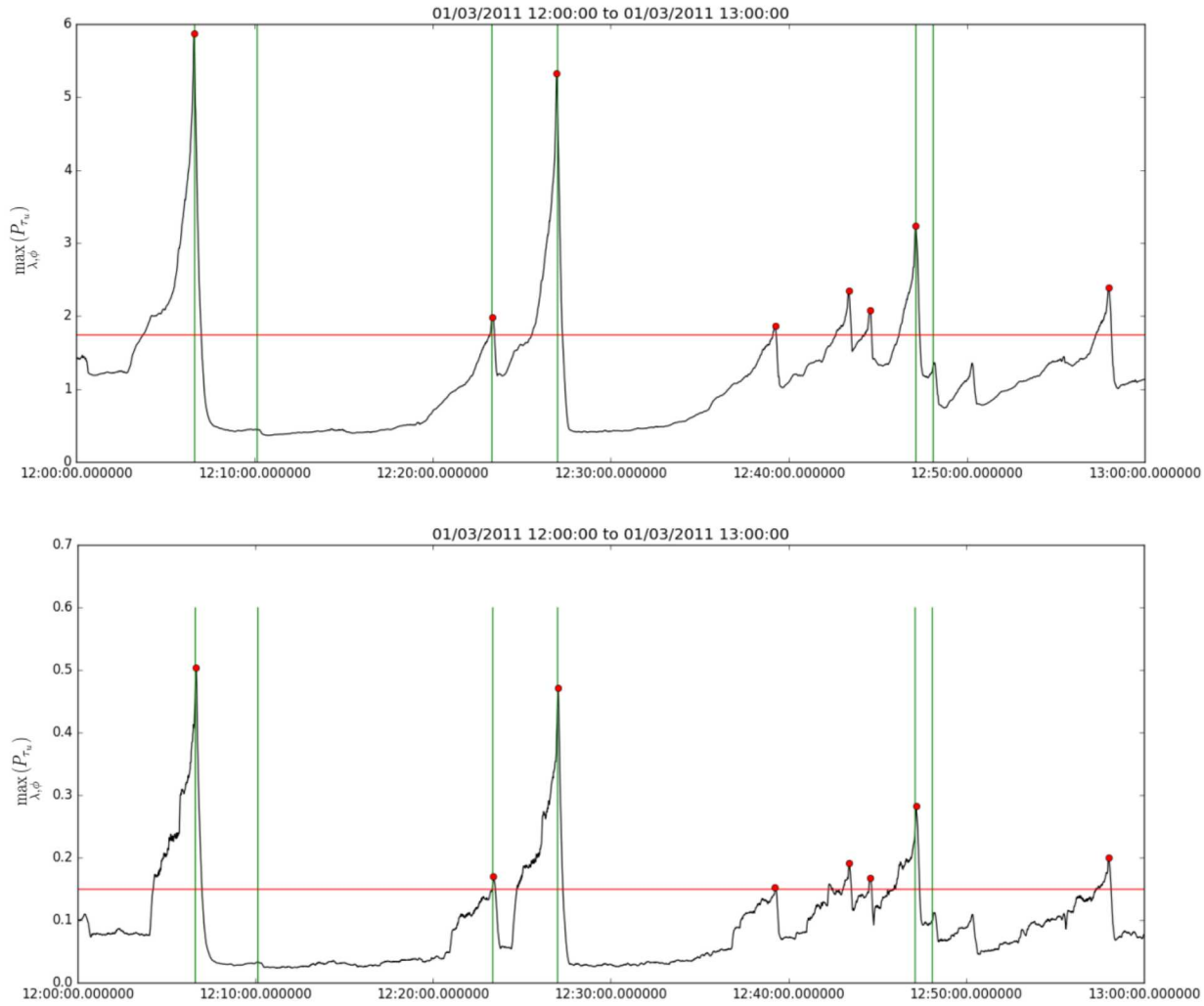

Figure 13. A comparison between the temporal variation of peak power, $\max_{\lambda,\phi}(P_{\tau_u})$, obtained using an empirical stack (top) and equivalent-resolution synthetic stack that does not include amplitude information (bottom). Red dots denote detections given the thresholds (red lines) in each panel. Vertical green lines denote the origin times of events in the UUSS earthquake catalog.