Evaluation of a pair-based, joint-likelihood association approach for regional infrasound event identification

Philip Blom, Garrett Euler, Omar Marcillo and Fransiska Dannemann Dugick

Los Alamos National Laboratory, Earth & Environmental Science Division, Los Alamos, NM 87545, USA. E-mail: pblom@lanl.gov

Received 2020 February 13; in original form 2019 August 19

SUMMARY
A Bayesian framework for the association of infrasonic detections is presented and evaluated for analysis at regional propagation scales. A pair-based, joint-likelihood association approach is developed that identifies events by computing the probability that individual detection pairs are attributable to a hypothetical common source and applying hierarchical clustering to identify events from the pair-based analysis. The framework is based on a Bayesian formulation introduced for infrasonic source localization and utilizes the propagation models developed for that application with modifications to improve the numerical efficiency of the analysis. Clustering analysis is completed using hierarchical analysis via weighted linkage for a non-Euclidean distance matrix defined by the negative log-joint-likelihood values. The method is evaluated using regional synthetic data with propagation distances of hundreds of kilometres in order to study the sensitivity of the method to uncertainties and errors in backazimuth and time of arrival. The method is found to be robust and stable for typical uncertainties, able to effectively distinguish noise detections within the data set from those in events, and can be made numerically efficient due to its ease of parallelization.

Key words: Probability distributions; Statistical methods; Acoustic properties; Earthquake monitoring and test-ban treaty verification; Wave propagation.

1 INTRODUCTION
Subaudible acoustic signals, termed infrasound, propagate through waveguides in the middle- and upper atmosphere that allow frequent observation at distances of hundreds to thousands of kilometres from the originating source, and are therefore of interest to a number of applications including detonation detection, natural hazard monitoring, atmospheric sounding and others. Bayesian frameworks for localization and characterization of infrasonic sources have been shown to be robust methods for analysis due to the need to quantify uncertainty of propagation effects produced by the dynamic and poorly sampled middle- and upper atmosphere through which infrasonic signals propagate (Blom et al. 2015, 2018). A similar Bayesian framework can be developed for infrasonic event identification through detection association using some of the tools previously developed for localization analysis.

The task of association in network level seismoacoustic signal analysis requires the identification of subsets of detections generated by a common source from within a larger set of detections collected across a spatially distributed network of stations. Such analysis, often termed event identification, is required for continued investigation of the source (localization, characterization, etc.) and is therefore of great interest for seismoacoustic researchers and analysts. Array- and network-level signal analysis algorithms are based on a combination of a statistical framework to quantify confidence in results and a physical model describing the propagation of signals to the network stations. Unlike the relatively simpler case of array-scale analysis for which propagation can be reduced to that of a planar or spherical wave, acoustic propagation between network stations is significantly more complicated and requires a thorough understanding of the propagation of infrasonic energy through the middle- and upper atmosphere.

While association of detections in the creation of an event bulletin or catalogue is a recent endeavor in the domain of infrasound, it has a long history in seismology. For a large portion of the 20th century, before the use of digital networks, association of seismic detections was a tedious, manual task (Richter 1958). Scientists would focus on reporting individual station bulletins of S & P times as well as any additional phases that were readily identifiable (while hopefully not forcing phase misidentification). The particle motion of P and surface waves were also sometimes used to constrain the azimuth to an event. Through the use of meticulously compiled time–distance charts (often validated with ground truth from large explosions), these sets of detections from a single station were converted into distances and origin times. This bulletin would then be compiled with other station bulletins and once absolute timing issues were resolved, a string, a large globe and a protractor were used on the detection sets with similar origin times to associate them (or not)
to a common region on the Earth’s surface. With this an event hypothesis was found for subsequent refinement. Because of the time-consuming manual approach and exchange of information, the process to produce a final bulletin of events might take around a year. To reduce this delay, seismologists tended to limit their catalogues substantially by focusing on well-recorded events.

Around the mid-1960s, seismologists began to look at automating the task of catalogue formation for rapid monitoring of earthquakes and nuclear explosions. For example, Seipp et al. (1968), under the Defense Advanced Research Projects Agency (DARPA) Project Vela Uniform was the first to devise such a system for a regional digital network. Unfortunately, this effort had limited impact on the field of seismology at the time as digital systems were not widespread. The advent of widespread digital seismic networks and communication infrastructure in the 1970s (Dorin & Eastlake 1978) brought about renewed interest in the operational problem of association for earthquake (Anderson 1978; Lee & Stewart 1981; Allen 1982) and nuclear explosion (Myykelveit & Bungum 1984; Blandford & Gomez 1985; Kerr 1985) monitoring. A key difference between these two is that the nuclear explosion monitoring community relied primarily on a select few well-tuned arrays that provide additional information (e.g., slowness) through array analysis that supports improved a priori phase identification and spatiotemporal constraint in comparison to the earthquake monitoring community that uses only arrival time data from individual sensors.

New developments continued into the 1990s, with refinements to both earthquake early warning and nuclear explosion monitoring association algorithms for large networks at local (Johnson et al. 1994) to global scales (Le Bras et al. 1994a, b; Engdahl et al. 1998), in some cases using artificial intelligence techniques (Bache et al. 1990; Beall et al. 1991; Bache et al. 1993). The Preparatory International Data Center (PIDC) and subsequently the International Data Center (IDC) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) adopted global association (GA) (Le Bras et al. 1994a, b) for their network processing of seismic, infrasound and hydroacoustic data (Brown et al. 2002). Currently, the IDC is considering moving to a Bayesian pipeline approach (Arora et al. 2013) that can consider null results (the lack of detection) in network processing. Recently, another probabilistic approach (Draelos et al. 2012, 2015) has been independently investigated that is quite successful with aftershock sequences and the USGS has recently revised their association approach (Kuyuk et al. 2014; Patton et al. 2016; Yeck et al. 2019). Current association research in seismology has also seen a revitalization with techniques for teleseismic event processing with regional networks (Jin et al. 2015), incorporation of machine learning algorithms (Ross et al. 2019; McBrearty et al. 2019) and waveform-based approaches (Arrowsmith et al. 2016; Bergen & Beroza 2019; McBrearty et al. 2019) to reduce analyst burden in catalogue creation.

Early infrasound research focused on automation of station-level analysis (Kerr 1971) while network-based catalogue creation in infrasound remained largely a manual task (Whitaker, personal communication). Automated association research with infrasound networks began with the creation of the International Monitoring System (IMS) by the CTBTO (Brown et al. 2002). In infrasound analysis only arrays are used to differentiate infrasound signals from local meteorological conditions (Brown 1963) unlike in seismology-based nuclear explosion monitoring, where both arrays and triaxial stations contribute to the association task. The current approach to infrasound processing at the IDC, called Progressive MultiChannel Correlation (PMCC) (Cansi 1995; Le Pichon et al. 2008), uses multi-band coherence that has been significantly tuned over the last decade. Regional approaches to infrasound association have also been developed (e.g. Arrowsmith et al. 2008) and interest in combining infrasound and seismic networks is growing (Stump et al. 2002; Hagerty et al. 2002; Arrowsmith et al. 2010; Gibbons et al. 2015). Recently, Arrowsmith et al. (2015) introduced a gridless approach based on a binary pair-wise measure of association. Such gridless approaches potentially reduce the computational burden over gridded approaches and are naturally adaptive to a dynamic network.

The discussion presented here is organized as follows. A gridless, pair-based association approach is introduced and detailed for use in infrasonic signal analysis. The association measure is then defined using a Bayesian formulation for the joint-likelihood of detection pairs such that the propagation physics is contained within the likelihood function. An overview of clustering analysis using hierarchical agglomeration is provided followed by a discussion of how clusters are defined via a linkage threshold and a quality control refinement is introduced considering cluster shape. A synthetic data set is utilized to evaluate the performance of the method on a sparse, regional scale network for which network stations are separated by hundreds of kilometres. Finally, the sensitivity of this association approach to errors and uncertainties in the detection characteristics, notably the arrival time and direction of arrival, are investigated.

2 A PAIR-BASED ASSOCIATION APPROACH

Many association frameworks utilize a coarse localization analysis to identify sets of detections potentially originating from locations in a spatial and temporal grid of hypothetical sources. An alternate approach for identifying events within a large set of detections is to analyse individual pairs of detections using a robust propagation model to determine whether they are potentially from a common source and identify those groupings where all members have high levels of association with others in the group. This allows the propagation physics to drive the associations and removes the dependence of the method on grid resolution. A pair-based association method was previously investigated by Arrowsmith et al. (2015) with promising results; however, that approach defined association using a binary formulation in which a given detection pair was either associated with a possible common source or unassociated. Such an approach simplifies the identification of clusters into a straightforward task using graph theory as shown in the left panel of Fig. 1. In the figure, each coloured circle represents a detected infrasound signal and the linkages represent decisions made that those pairs are consistent with a common source. The blue and red clusters denote two identified events from the detection set, while the grey circles are unassociated with other detections and could be attributed to local sources not observed by other observers or possibly as noise/false detections. The line connecting a single detection in the blue event with a single grey detection indicates a case in which a local noise detection might be consistent with one detection within an event, but not others. In the second event, one of the connections is not determined to be a valid association. These and a number of other complications including the addition of ad hoc rules requiring large numbers of detections in events limit the performance of this approach for sparse networks as discussed by Arrowsmith et al. (2015).

A more robust and versatile scheme for association analysis can be developed using an association quantification defined by some normalized scalar measure and a hierarchical clustering algorithm...
to identify events using a dendrogram as shown in the right panel of Fig. 1. The joint-likelihood for each pair defines this scalar association measure, \( A_{ij,k} \), and a separation distance computed from the negative logarithm of the joint-likelihood, \( d_{i,j} = -\log(A_{ij,k}) \), is used in a hierarchical clustering analysis. This analysis uses a threshold value (the horizontal dashed line in the figure) to define a cut-off and determine the minimum accepted separation level. This threshold level provides a means to cut-off local or noise detections incorrectly associated with subsets of detections within event clusters and similar complications. The spurious connection between one element in the blue event and a noise detection in the left-hand panel of the figure is now represented in the dendrogram as the link between detection 7 and the cluster of detections 8, 9 and 10. This linkage is found to occur at a higher point along the separation scale, so that the choice of threshold value prevents its inclusion in the event. Similarly, the missing connection between a pair of elements in the red event is found to occur just below the threshold so that detection 3 is included in the cluster with detections 4, 5 and 6. In addition to the minimum association level threshold, identifying events requires a minimum population of detections and a hierarchical clustering linkage approach to determine the clustering of detections. The linkage threshold and population limit can be physically or empirically defined and provide a simple and straightforward means of tuning the algorithm for a given network and a means to generalize the pair-based association approach to other phenomenologies or combinations of phenomenologies.

Previous analysis has found that infrasonic localization algorithms based on Bayesian approaches provide accurate estimates of source localization and uncertainty quantifications (Modrak et al. 2010; Marcillo et al. 2014; Morton & Arrowsmith 2014; Blom et al. 2015). A similar approach can be applied to the problem of associating events. This problem requires a minimum population of detections and a hierarchical clustering linkage approach to determine the clustering of detections. The linkage threshold and population limit can be physically or empirically defined and provide a simple and straightforward means of tuning the algorithm for a given network and a means to generalize the pair-based association approach to other phenomenologies or combinations of phenomenologies.

Figure 1. Cluster-based analysis of binary linkages (left-hand panel) and hierarchical analysis of linkages (right-hand panel) for pair-based association. The separation measure for hierarchical analysis, \(-\log(A_{ij,k})\), is defined from the joint-likelihood for each detection pair as in eq. (2). The dashed line in the right-hand panel denotes a clustering threshold below which linkages are accepted to form the same two groups of detections as the binary linkage approach presented in the left-hand panel.

Thus, the pair-based, joint-likelihood association analysis proposed here first requires computation of \( A_{ij,k} = P(D_1|S) P(D_2|S) P(S) \) for each possible pair of detections considered in analysis. The efficiency of the method is limited by the time needed to numerically compute this integral for the various detection pairs, which is a primary concern in constructing the likelihood function definition. For a list of \( N \) detections, the number of unique pairs to consider is \( \binom{N}{2} \), which can be computationally intensive for large \( N \). Fortunately, the computation for each pair can be completed independently, so that the computation is easily parallelized.

2.1 Defining the source, detection and likelihood

The source hypothesis, detection and likelihood definitions used in this analysis have been leveraged from the Bayesian Infrasonic Source Localization (BISL) methodologies (Modrak et al. 2010; Blom et al. 2015). At leading order, the source hypothesis can be defined in terms of only the temporal and spatial description of the source,

\[
S = \{x_s, y_s, \tau_s\},
\]

(3a)

where \( x_s \) and \( y_s \) denote the spatial location and \( \tau_s \) is the event time. Unlike many seismic association and localization methods, the depth (or elevation) of the source is not included in this hypothesis. This is because signals produced by underground sources such as earthquakes and explosions are actually generated by the motion of the ground surface so that the infrasonic source is located at the ground in such cases. In the case of an elevated source, the infrasonic propagation statistics are essentially independent of the source altitude so that regional to global distance observations from surface and elevated source are mostly indistinguishable. It is possible to include other features that may constrain additional source characteristics such as peak overpressure, spectral structure, or duration; however, for this initial investigation, only the temporal and spatial coincidence of hypothetical sources will be considered.

The quantitative information gained from a detection is assumed to include the array location, detection time, and beamforming results,

\[
D_j = \{x_j, y_j, \tau_j, \varphi_j, \mathcal{F}_j\}.
\]

(3b)
where \( x_i \) and \( y_i \) are the spatial location of the detecting array, \( \tau_j \) is the time of the detection, \( \psi_j \) is the backazimuth obtained from beamforming analysis, and \( F_j \) is the Fisher statistic that can be related to the coherence of the signal across the array and is used to define the weighting between detections. Once again, additional characteristics of the source may be utilized in refining the association measure between pairs which would require including observed overpressure, spectral structure, or other information in the detection; however, such inclusions require more robust propagation models and complicate evaluation of the joint-likelihood integration so that, for this initial investigation of the method, only spatial and temporal coincidence is considered here.

Following the development used in the BISL framework, the likelihood function can be separated into the product of azimuthal dependent and time-range dependent contributions (Blom et al. 2015). Infrasonic propagation paths extending into the middle- and upper atmosphere can interact with strong cross winds resulting in significant deviation between the observed backazimuth and that of the true source. Further, the exact influence of the cross winds are variable and uncertain due to the dynamic and poorly sampled nature of the atmosphere. Because of this uncertainty and the finite range of possible backazimuth values, the von Mises distribution commonly utilized in directional statistics has been leveraged to define the azimuthal dependence of the likelihood (Mardia & Jupp 2009). The time-range dependent portion of the likelihood is defined by a distribution, \( \rho_r (r, \tau_j) \), which quantifies the possible infrasonic celerities (horizontal group velocities) for infrasonic propagation. The likelihood of a given source producing the observed detection can then be defined from the product of the von Mises distribution with \( \rho_r \),

\[
P(D_j | S) = \frac{e^{\kappa \cos(\phi - \psi_j)}}{2\pi I_0(\kappa)} \rho_r (r, \tau_j - \tau_j),
\]

where \( \phi \) is the azimuthal direction between the detecting station and hypothetical source, \( r \) is the source-station range and \( I_0(\kappa) \) is the modified Bessel function that normalizes the von Mises distribution. It should be noted that although \( x_s, y_s \) and \( x_r, y_r \) in the above definitions denote Cartesian coordinates, the source and station locations can easily be generalized for latitude/longitude definitions because the likelihood function leverages the station-source azimuth, \( \phi \), and separation, \( r \), which can be computed using Cartesian coordinates or the great circle bearing and path length.

The von Mises width, \( \kappa \), is defined using an updated analysis from that in Blom et al. (2015) which directly relates the Fisher statistic to the Capon beamwidth using the signal and noise eigenvalues, \( \lambda_s \) and \( \lambda_n \), respectively, and number of sensors in the array, \( N \). Assuming a single incident plane wave on the array, an eigenvalue decomposition of the data covariance matrix produces a set of eigenvalues, \( \lambda_s, \lambda_s, \ldots, \lambda_s \), which leads to,

\[
F = \frac{P_{\text{bm}}}{P_{\text{total}} - P_{\text{bm}}} \frac{(N - 1)}{\lambda_s}
\approx \frac{\lambda_s}{(\lambda_s + (N - 1) \lambda_n) - \lambda_s} (N - 1) = \frac{\lambda_s}{\lambda_n},
\]

which can be combined with the curvatures of the Capon beam and von Mises distribution at the azimuth of the peak to yield (Blom et al. 2015),

\[
\kappa = 2(N - 1) F.
\]

This beam uncertainty is combined with a fixed uncertainty due to propagation effects, \( \Delta = 4^\circ \), to produce a final definition of \( \kappa \) for use in the likelihood,

\[
\kappa = \frac{\ln(2)}{1 - \cos \left( \phi_1 + \Delta \right)}.
\]

where the beaming analysis halfwidth is defined from the \( F \) value,

\[
\phi_1 = \cos^{-1} \left( 1 - \frac{\ln(2)}{2(N - 1) F} \right).
\]

The value of \( \Delta \) has been chosen based on propagation statistics for the western United States previously analysed to define regional propagation models (Blom et al. 2015). In practice, one identifies the \( F \) value of an arrival from detection analysis and uses it to define the value of \( \phi_1 \), which is then combined with the propagation uncertainty to define \( \kappa \) for use in computing the likelihood.

### 2.2 Propagation models

Propagation-based, stochastic models for path geometry and travel-times have been introduced and studied in applications of infrasonic localization (Blom et al. 2015); however, for the application of pair-based association to networks with high detection volumes, a simple generalized propagation model is preferred for its efficiency in evaluating the potentially large number of detection pairs. A generalized, 1-D celerity model can be constructed by analysing the distribution of predicted arrival celerity for a large number of atmospheric specifications without separating out seasonal variations and arrival location. Following the construction of propagation-based, stochastic path geometry models in Blom et al. (2015), ray paths have been generated using inclinations covering \( 1^\circ \) to \( 50^\circ \) with steps of \( 1.5^\circ \) and azimuthal steps of \( 5^\circ \) for infrasonic signals propagating through atmosphere specifications for 2007 through 2013 as obtained from the Ground-to-Space (G2S) archive (Drob et al. 2003, 2010). Atmospheric absorption has been computed using the Sutherland & Bass (2004) formulation and arrivals with predicted attenuation (combined geometric spreading and absorption at 1 Hz) in excess of 40 dB are dropped from the analysis. The resulting distribution of celerity values can be approximated using a kernel density estimate (KDE) of the scatter of celerity values and is shown in the blue curve of Fig. 2.

In order to further improve the numerical efficiency of the method, the celerity model utilized for association analysis has been parametrized as a Gaussian mixture model (GMM) defined in terms of inverse celerity, \( v^{-1} = \gamma \), so that the propagation time is in the numerator of the exponential argument and the corresponding component of the integration can be completed analytically using...
Table 1. Means, \( \mu_j \), variances, \( \sigma_j \), and weights, \( w_j \), for the Gaussian mixture model (GMM) defining infrasonic celerity as shown in Fig. 2.

<table>
<thead>
<tr>
<th>( j )</th>
<th>( \mu_j ) (skm(^{-1}))</th>
<th>( \sigma_j ) (skm(^{-1}))</th>
<th>( w_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.327</td>
<td>0.066</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>0.293</td>
<td>0.080</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>0.260</td>
<td>0.330</td>
</tr>
</tbody>
</table>

\[(\text{Bromiley 2003}),\]

\[
\int \frac{1}{\sqrt{2\pi \sigma_1^2}} e^{-\frac{1}{2} \left(\frac{x-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sqrt{2\pi \sigma_2^2}} e^{-\frac{1}{2} \left(\frac{x-\mu_2}{\sigma_2}\right)^2} \, dx = \frac{1}{\sqrt{2\pi \left(\sigma_1^2 + \sigma_2^2\right)}} e^{-\frac{1}{2} \left(\frac{x-\mu_1-\mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)^2}. \quad (6)\]

A three component mixture model has been used to parametrize the distribution constructed in the KDE in Fig. 2 and is shown as the red line in the figure. The parameters defining the GMM are summarized in Table 1 and the individual components are found to correspond to returns from the tropopause, stratopause, and thermosphere, with central celerities of 327, 293 and 260 m s\(^{-1}\), respectively. This parametrization produces an accurate representation of the KDE distribution except for a slight difference in the low celerity tail; though, this difference is expected to be insignificant in analysis. Using this simplification, the multivariate integration required to calculate the association for a pair of detections reduces to a 2-D numerical integral over a spatial region of interest where the projected backazimuths intersect.

2.3 Defining the integration region and evaluating the integral

Efficient computation of the joint-likelihood in eq. (2) first requires identification of the region of the source hypothesis space over which \( P(D_1|S)P(D_2|S)P(S) \) is non-zero and the integral should be evaluated. The inverse celerity formulation used in the propagation model enables the source time integration to be evaluated analytically, so that defining the integration region for the numerical computation reduces to identifying a spatial area for integration from direction of arrival information. For the analysis here, the integration area is defined by a central point and radius from which the source could be located. In the case that both detections are from the same array, the range- and azimuth dependence of the integration separate and the association value is easily computed as the product of two 1-D integrals over a region with some limiting range, \( R_{\text{max}} \), centred at the detecting array. However, in the more general case for which the detections are on different arrays, the region must be identified using the intersections of the projected backazimuths of the individual detections.

The integration region can be identified using the intersections of the primary direction of arrival beams and bounding beams defined by some fixed width as shown in Fig. 3. In the figure, the dark and light blue lines denote the primary and bounding beams for each detection, respectively, and the nine points are located at the resulting intersection points. The green point denotes the intersection of the primary beams, the dark blue points denote those intersections involving a combination of primary and bounding beams, and the light blue points denote those of only the bounding beams. Note that the case shown in the figure is idealized and some of the intersection points may not exist for a given detection pair.

The analysis of these intersection points to define the integration region is as follows. Firstly, backazimuths are projected from each detecting array to some maximum propagation range, \( R_{\text{max}} \), and the resulting intersection points are identified. In the case that none of the primary intersection points (green and dark blue in Fig. 3) exist, then the association is set to zero, otherwise the geographic mean of all primary intersection points is defined as the central point for the integration region. In the case that neither array is within the projected beam of the other, the integration region radius is easily defined from the average distance between the defined central point and all secondary intersection points (that is, the dark- and light blue points in the figure), which identifies the region within the green circle in Fig. 3 as the integration region. An example of such a solution is shown in the leftmost panel of Fig. 4. For more complicated cases in which one or both arrays are within the projected beam of the other, the radius of the region extends from the central point to the nearer of the arrays as shown in the middle and right panels of Fig. 4. Note that, in the case of a pair of arrays both within the beam of the other, as in the right panel of the figure, this solution results in a large integration region.

Several quality checks are then made on this region definition to ensure the integral can accurately estimate the joint-likelihood. The radius is compared with minimum and maximum values, \( r_{\text{min}} \) and \( r_{\text{max}} \), and modified accordingly and if either array is contained within the integration region, the central point is shifted along the great circle path of that detecting array’s beam so that it is no longer contained within the integration region. Finally, once the region is defined, the joint-likelihood can be integrated numerically to obtain the association level between the two detections. For the investigation and evaluation performed here considering regional distance propagation, a width of 7.5° is used so that the total angle between bounding beams is 15°, a maximum range of 2500 km has been chosen, and the minimum and maximum integration region radii are set to 100 and 1000 km, respectively, though different values may be appropriate in the case of global scale analysis.
2.4 Clustering the joint-likelihood results

Identification of events within a set of detections given their joint-likelihood values can be accomplished using a hierarchical agglomerative analysis (Kaufman & Rousseeuw 1990) to cluster those subsets with mutually high association. Visualization and analysis of such data is typically performed using a distance matrix in which the diagonal elements are set to zero and the off diagonal elements are defined by a measure of Euclidean or non-Euclidean distance between elements. Again, this separation is related to the joint-likelihood by a negative-logarithm, \( d_{i,j} = -\log_{10}(A_{i,j}) \), and the resulting distance measure is non-Euclidean as the individual detections are not defined by specific points in a vector space but only by the overlap of their likelihood functions. An example of such a distance matrix is shown in Fig. 5 for a data set containing three events and a number of noise detections. The colour scale shows the distances for the various pairs and, because the horizontal indices are identical the vertical ones, they are suppressed for brevity.

As mentioned previously, the identification of integration regions occasionally finds a number of pairs that have no overlap in their direction of arrival projections so that their joint-likelihood values are set to zero. Such results are converted to logarithmic space by defining a maximum separation to avoid an infinite separation value that causes instabilities in the clustering analysis. For the analysis here, an upper bound of 8, corresponding to a joint-likelihood value of \( 10^{-8} \), has been found to provide stability of the clustering analysis.

Clustering analysis of the distance matrix is performed using linkages defined by the Weighted Pair Group Method with Arithmetic Mean (WPGMA) (Sokal & Michener 1958). This analysis produces a dendrogram that sorts and merges the detections into clusters dependent on their separations in the distance matrix as shown in the left sides of the panels in Fig. 6. In the figure, the indices on the vertical axis of the dendrograms denotes the detection IDs, while the horizontal axes show the linkage distance. The data used in generating these examples are one realization from a synthetic data set that will be discussed in more detail in Section 3. In this realization, detections 0–8, 9–15 and 16–19 each correspond to a unique event, while detections 20–27 are not associated with any events and have been added as random noise to the data set. On the right-hand side of the panels are the sorted distance matrix previously shown in Fig. 5 where the ordering of detections had been randomized to visually obscure the three events. Note that due to the detection ID sorting within each defined cluster, the detection order on each dendrogram and sorted distance matrix differ.

Identification of clusters from the dendrogram results is completed by defining a threshold separation value, \( \alpha_{\text{linkage}} \), denoted as the vertical dashed line in each of the dendrogram panels. For any given threshold value, events are defined by those sets of detections linked below the threshold, and the resulting groupings of detections are denoted by the shaded boxes in the sorted distance matrix in Fig. 6 and the colour of linkage lines in the dendrogram. For the analysis here, a minimum dimension of 3 detections on 2 stations are required to be below the linkage threshold, \( \alpha_{\text{linkage}} \), in order for a cluster to be determined to be an event so
Figure 6. Clustering analysis result without (upper panel) and with (lower panel) trimming for an example data set in which detections 0–8, 9–15 and 16–19 correspond to three distinct events. The dendrograms (left of each panel) and sorted distance matrix (right of each panel) show the structure of the detection linkages and the grouping of detections into event clusters, respectively. The colour scales for the distance matrices correspond to that in Fig. 5, and the vertical dashed line denotes the clustering threshold, $\alpha_{\text{linkage}} = 5.5$, below which linkages are accepted. The shaded boxes in the distance matrices denote the identified event clusters. The X in the distance matrix of the lower panel denote a linkage that have been ‘trimmed’ due to poor cluster shapes as discussed in the text.
that localization can be feasibly completed. For the upper panel, this produces 4 sets of detections defining event clusters denoted by the blue, magenta, red, and green sets of detections in the dendrogram.

In order to correctly identify the event containing detections 0–8, the linkage threshold, $\alpha_{\text{linkage}}$, had to be set to an overly high value of 5.5 in the upper panel of Fig. 6 due to a spurious linkage between detections 5 and 26. This is a common complication in clustering analysis that is due to a strong association between a single member of an event with a non-member. This spurious association is likely produced by a scenario such as that shown in Fig. 7. The blue arrows in the figure denote backazimuth projections for detections produced by an event, while the single grey line denotes a spurious detection that may be incorrectly included in the cluster should the joint-likelihood between it and the detection from the upper right station be sufficiently large. In such a case, the dendrogram linkage between the detection on the upper right station with the spurious detection is made prior to its linkage with other detections from the event, which either pulls the spurious detection into the cluster or pulls the detection from the upper right station out of the cluster depending on the choice of linkage threshold, $\alpha_{\text{linkage}}$.

A means of identifying and correcting such spurious linkages can be developed by first defining a cluster shape parameter,

$$\gamma = \frac{d_{\text{max}}}{\bar{d}},$$

(7)

where $d_{\text{max}}$ is the maximum distance between any pair of detections in the cluster (which defines the cluster diameter) and $\bar{d}$ denotes the median interelement distance for all members of the cluster. Consider a cluster defined by a submatrix within the original distance matrix defined by,

$$D = \begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0.00 & 2.10 & 1.91 & 5.48 & 1.93 & \vdots \\
2.10 & 0.00 & 1.93 & 8.00 & 2.93 & \vdots \\
1.91 & 1.93 & 0.00 & 6.85 & 1.97 & \vdots \\
5.48 & 8.00 & 6.85 & 0.00 & 1.29 & \vdots \\
1.93 & 2.93 & 1.97 & 1.29 & 0.00 & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots 
\end{pmatrix}$$

The median inter-element spacing in this cluster is $\bar{d} = 2.04$ and the cluster diameter is $d_{\text{max}} = 8.0$, so that the shape parameter, $\gamma = 3.93$. A parametric study of the shaping parameter with a number of scenarios has been undertaken and a threshold, $\alpha_{\text{shape}} = 3.8$, has been identified such that, if $\gamma > \alpha_{\text{shape}}$, trimming improves the clustering results without impacting true associations at regional distances. As with other parameters of this analysis, the value of $\alpha_{\text{shape}}$ may vary in certain conditions such as global scale propagation distances.

Once it has been determined that a cluster needs to be trimmed, identifying the linkage causing the problem is straightforward. For each element of the cluster, the average spacing can be defined by its distance to other detections within the cluster,

$$\tilde{d}_j = \frac{1}{K-1} \sum_{k \neq j} d_{j,k},$$

(8)

The element of the cluster for which this value is largest is the first index of the linkage that should be trimmed, and the second index can be identified by the minimum separation between this first detection and others in the cluster. For the above matrix example, this results in identification of the row containing [5.48 8.00 6.85 0.00 1.29] as defining the first index and the smallest value linking this with another detection is the 1.29 value that defines the linkage’s second index. This linkage distance can be replaced by the maximum distance (8), and the modified distance matrix is then clustered to identify the new results. For this example, trimming reduces the cluster shape parameter from 3.93 to 1.50. In general, spurious linkages can be identified and trimmed iteratively until none of the clusters identified in the distance matrix have poor shape parameters. Finally, trimming produces a new cluster containing one fewer detections than the original, so any clusters already containing the minimum population of detections that are identified as poorly shaped clusters are thrown out and ignored.

Applying this trimming method to the results in the upper panel of Fig. 6 produces the result in the lower panel of the figure. The spurious linkage between detections 5 and 26 has been trimmed and is denoted by the “X” in the distance matrix and the expected clustering of detections 0–8, 9–15 and 16–19 is 15 is produced. In the trimmed results, the clustering threshold $\alpha_{\text{linkage}}$ can be lowered to 3.5, which removes the cluster defining the magenta event in the initial analysis and produces the expected result of three events formed by detections 0–19. Reducing the threshold post trimming is difficult to do in an automated manner, so that the analysis used to investigate the performance and sensitivity of the association algorithm, $\alpha_{\text{linkage}}$ does not include a post-trimming lowered threshold which may result in noise detections being included in clusters unnecessarily. Thus, the results in the following sections provide limiting bounds on performance and sensitivity of the algorithm as

Figure 7. A backazimuth based example of a poorly shaped cluster. The four dark-blue arrows denote backazimuth projections from detections produced by a single event, while the grey arrow denotes a spurious detection that could be pulled into the cluster should its association be sufficiently strong with the detection on the upper right station by their backazimuth intersection in the upper right.
an automated method that would be improved by an analyst review and correction of clustering results.

An overview of the parameters used in the association algorithm is provided in Table 2 and shows the relative simplicity of the method. A number of the parameters, such as those defining the likelihood functions, the maximum distance matrix value, and the criterion for declaring a cluster to be an event, are not expected to be modified in most cases, while others including the $R_{\text{max}}$, $r_{\text{min}}$ and $r_{\text{max}}$ values have set values corresponding to a given propagation scale. Analysis of additional data sets of varying propagation scale and ideally with ground truth sources to determine performance is required to more thoroughly tune these parameters for use.

### 3 EVALUATION OF THE ASSOCIATION METHOD

In order to evaluate the performance of the pair-based, joint-likelihood association approach detailed here, a synthetic data set has been constructed using a network of arrays and several potential source locations in the western United States. The data set utilizes synthetic detections on four infrasonic arrays including regional arrays in New Mexico, Wyoming and Nevada as well as an infrasonic array in California operated as part of the IMS as shown in Fig. 8. For simplicity in construction of the synthetic data set, each detection is defined an $F$ value corresponding to a halfwidth of $\phi_1 = 7.5^\circ$ in eq. (5c) (for a 5-element array, this corresponds to an $F$ value of approximately 10). Infrasonic source locations have been selected to include a pair of locations contained within the network at the Nevada National Security Site (NNSS) and Utah Test and Training Range (UTTR) as well as one location near the outer edges of the network at White Sands Missile Range (WSMR) where the azimuthal coverage is poor. Locations of the network arrays and locations and origin times for the sources used in the synthetic data set are summarized in Table 3. It should be noted that the quantitative results in this and the following discussions are unique to this specific network geometry, the sources used in the analyses, and the modelled propagation effects; however, qualitative assessments can be gleaned from these results that are broadly applicable to other network designs, source locations and propagation scenarios.

Propagation paths have been computed using a G2S profile for the western United States on 1 January 2010 in order to define realistic detection information for evaluating the association algorithm. Propagation paths connecting these specific source–receiver geometries, termed eigenrays, have been computed assuming a stratified, moving atmosphere using the approach developed by Blom & Waxler (2017) to produce the detections summarized in Table 4. The backazimuth projections for these detections separated by individual events are shown in Fig. 9. The presence of cross winds in the atmosphere produces biases in the backazimuths of observations so that back projecting the arrivals does not lead to a single intersection, but instead produces a number of distinct intersection points near the source location. The scale and direction of these biases are noted in Table 4 using the convention that the bias is the observed azimuth minus that of the actual source location. It should be noted that source energies are not considered in this analysis and a number of the propagation paths, particularly those extending into the thermosphere, may not correspond to observable signals for small amplitude sources. As in the construction of propagation statistics, those propagation paths with excessive attenuation due to geometric and absorption losses have been removed as have many of the multi-pathed tropospheric phases. For a more robust analysis of a full signal analysis pipeline, waveform predictions could be computed and combined with a noise model to produce detection statistics; however, such analysis is beyond this initial investigation of the association algorithm. Regardless, this synthetic data in Table 4 provides a realistic set of propagation times and backazimuth deviations to evaluate this or other association and localization algorithms.

In order to create a realistic synthetic data set, noise detections have been introduced in the analysis of the data set. Two additional detections have been generated for each array with randomized backazimuth and arrival times. The backazimuth values are generated randomly from $-180^\circ$ to $180^\circ$ and the arrival time values have been generated using a uniform distribution from 12:00:00 to 14:30:00 to be consistent with the source time and propagation times. In order to ensure that the noise detections are inconsistent with the sources in Table 3, the likelihood function for each noise detection is evaluated at the source locations and origin times with the condition that the detection be regenerated if the probability of the source producing that detection is larger than 1e-10. One such set of noise detections are included in Table 4 and have been used in generating the example clustering solution shown in Fig. 6 where the green, red, and blue clusters correspond to the events at NNSS, UTTR, and WSMR, respectively. This colour mapping has also been used in Fig. 9 and will be used in subsequent figures.

The performance of the association algorithm has been evaluated using 100 randomly generated sets of noise detections and the resulting clustering solutions are shown in Fig. 10. Two scenarios have been considered to more thoroughly evaluate the method. First, only the sources at UTTR and WSMR have been included in the analysis in order to estimate the performance when sources are spatially separated and easily resolved by the detection backazimuths. The leftmost panel of the figure shows the resulting cluster labels for the 100 noise samples and the first two sets of data in the histogram on the right show the count of corrupted event clusters (when the detections defined to be in the event are not correctly included) and false associations (when noise detections are incorrectly included in the cluster). These counts are shown for each event separately and denoted by solid and dashed boxes, respectively.

For this case of spatially separate events at UTTR and WSMR, correct events are identified in more than 95 per cent of cases and noise detections are incorrectly included in clusters approximately 20 per cent of the time. The inclusion of noise detections is potentially due to the high shape parameter threshold of the automated trimming method utilized to optimize the correct event identifications and could be reduced by analyst review. The same 100 noise realizations were used to compute the clustering of detections when all three sources are included and the method is still able to correctly identify the UTTR and WSMR events more than 90 per cent of the time. The NNSS event is correctly identified in more than 85 per cent of cases, so that in general the pair-based joint-likelihood association method is able to correctly identify events on a regional network in 85–95 per cent of the cases considered. The false association frequency in the case of three events remains approximately constant at approximately 20 per cent, which is equivalent to the two-event case; though, there are only a few false associations for the NNSS event potentially due to the large number of detections in the event. In both the 2 and 3 event cases, the frequency of false detections is larger for the WSMR event than that at the UTTR.

Though not included in the figure, additional analysis has been conducted using the single event at NNSS with a larger noise data set of 5 non-event detections at each station. In these analyses, a number of scenarios were considered in which only a subset of the
Table 2. Summary of tunable parameters in the pair-based, joint-likelihood association algorithm.

<table>
<thead>
<tr>
<th>Analysis step</th>
<th>Parameter</th>
<th>Default value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood definition</td>
<td>$\Delta$</td>
<td>$4^\circ$</td>
<td>Propagation-related backazimuth uncertainty</td>
</tr>
<tr>
<td></td>
<td>$\mu$, $\sigma$, $\bar{w}$</td>
<td></td>
<td>Parameters defining the celerity distribution</td>
</tr>
<tr>
<td>Integration region ID</td>
<td>$\delta \phi$</td>
<td>$7.5^\circ$</td>
<td>Width of the bounding beams</td>
</tr>
<tr>
<td></td>
<td>$R_{\text{max}}$</td>
<td>$2500$ km</td>
<td>Distance to project primary and bounding beams</td>
</tr>
<tr>
<td></td>
<td>$r_{\text{min}}$</td>
<td>$100$ km</td>
<td>Minimum radius of the integration region</td>
</tr>
<tr>
<td></td>
<td>$r_{\text{max}}$</td>
<td>$1000$ km</td>
<td>Maximum radius of the integration region</td>
</tr>
<tr>
<td>Clustering analysis</td>
<td>$d_{\text{max}}$</td>
<td>$8.0$</td>
<td>Linkage threshold</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{\text{linkage}}$</td>
<td>$5.5$</td>
<td>Linkage threshold</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{\text{shape}}$</td>
<td>$3.8$</td>
<td>Cluster shape parameter threshold for trimming</td>
</tr>
<tr>
<td></td>
<td>$N_{\text{det}}$</td>
<td>$3$</td>
<td>Minimum number of detections to declare an event</td>
</tr>
<tr>
<td></td>
<td>$N_{\text{st}}$</td>
<td>$2$</td>
<td>Minimum number of stations to declare an event</td>
</tr>
</tbody>
</table>

Figure 8. Map of sources (red stars) and arrays (blue triangles) in the southwest United States used in constructing a synthetic data set to evaluate the association performance on a regional network.

Table 3. Network array locations and source locations and origin times used to generate synthetic data for evaluation of the association framework.

<table>
<thead>
<tr>
<th>Latitude</th>
<th>Longitude</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLIAR</td>
<td>35.857</td>
<td>-106.315</td>
</tr>
<tr>
<td>I57US</td>
<td>33.606</td>
<td>-116.453</td>
</tr>
<tr>
<td>NVIAR</td>
<td>38.430</td>
<td>-118.304</td>
</tr>
<tr>
<td>PDIAR</td>
<td>42.767</td>
<td>-109.594</td>
</tr>
<tr>
<td>NNSS</td>
<td>37.221</td>
<td>-116.061</td>
</tr>
<tr>
<td>UTTR</td>
<td>41.131</td>
<td>-112.896</td>
</tr>
<tr>
<td>WSMR</td>
<td>33.420</td>
<td>-106.428</td>
</tr>
</tbody>
</table>

NNSS detections in Table 4 were included in order to determine how fewer detecting stations and phases impacts identification of an event in more realistic scenarios. Association analysis was run with detections from 3 of the 4 listed stations as well as only 2 of the 4, and even in the limiting case of 3 detections at NVIAR and I57US, the event at NNSS is consistently identified in the clustering analysis. Thus, under the assumption that non-event detections can be distinguished using spatial and temporal differences, the joint-likelihood, pair-based association algorithm detailed here can consistently identify single events in noisy detection sets. A possible complication has been identified in this analysis related to the introduction of false alarms. Introducing 5 non-event detections at each station produced a number of 2- and 3-detection clusters in the analysis so that the population thresholds defined to limit event identification may need to be increased to 3 stations for realistic data set or, as mentioned previously, the likelihood definition may need to include additional information such as duration, signal amplitude, frequency band, or other features to distinguish detections from different sources.

4 SENSITIVITY ANALYSIS

4.1 Sensitivity to errors in detection backazimuth

Bias in the detected backazimuth relative to the true source direction can be introduced by propagation effects such as cross winds and horizontal gradients in the propagation medium as well as by limitations of beamforming analysis. In order to quantify the sensitivity of the association framework to such errors, a modified list of detections has been defined with the propagation-related backazimuth bias removed, resulting in the beams shown in Fig. 11. In this modified case every detection has a backazimuth identical to the source azimuth and random perturbations of a given scale can be applied to estimate the deterioration of the event definitions as errors in detection backazimuth increase.

In the following analysis, a single set of noise detections is included to remove any additional variability and isolate the sensitivity of the method to errors in detection backazimuth. The specific noise detections utilized in this analysis are those summarized in Table 4 that produce the clustering solutions in Fig. 6. Random perturbations to the backazimuth of each non-noise detection have been randomly generated from a uniform distribution on an interval $\pm \delta \phi$ resulting in 100 perturbed detection sets for analysis. These perturbations are only applied to the non-noise detections because the noise detections have been defined such that the likelihood of each originating from any of the sources is negligible and applying perturbations could produce a noise detection that no longer satisfies this assumption.

The resulting clustering results are summarized in Fig. 12 where the upper panels show the clustering labels obtained for only events at UTTR and WSMR included in the analysis. Once again, for spatially separated source locations, the algorithm is able to consistently
identify the two events and the inclusion of noise detections within the clusters increases slowly with increased backazimuth error. For reasonably realistic backazimuth errors up to $\pm 6^\circ$, the two events are correctly identified in every realization and noise detections are incorrectly included only when the azimuth deviations reach $8^\circ$. Even for large backazimuth errors on the order of $10^\circ$, the UTTR event is correctly identified in $\sim 80$ per cent of realizations.

Analysis of all three events produces increasing corrupted associations rates for the NNSS and UTTR events. The increased incorrect clustering solutions for these events is likely due to the network and source geometry. From Fig. 8, it is evident that for several of the arrays, notably PDIAR and I57US, the difference in backazimuth of the NNSS and UTTR source locations is small. Because of this, resolving the two separate events depends strongly on correctly identifying the source-station range, which is limited by the large uncertainty of the celerity model as shown in Fig. 2. In the case that the two events occurred at largely different origin times, the method would more easily resolve the two, but with similar origin times the method cannot correctly distinguish which detections are from which event. The degraded performance when considering large backazimuth errors for the combination of all three sources may be related to the difference in backazimuth of the two spatially similar sources for each of the arrays. The azimuthal difference between the two sources from DLIAR is $35^\circ$ and from NVIAR is $68^\circ$; however, the differences at I56US, I57US, and PDIAR are much smaller at $20.0^\circ$, $14.6^\circ$, and $13.5^\circ$, respectively. Once the deviations reach $8-10^\circ$, detections can be shifted by an amount so large that they come closer to the non-originating source azimuth and cluster with it instead of the original event. When two sources are located in such a way, it is possible that dropping PDIAR or I57US from this analysis due to their inability to resolve the two source azimuths may improve the identification of events from UTTR and NNSS.
Pair-based association for infrasound

Figure 10. Clustering results for one hundred realizations of noise detections using two spatially separated sources at UTTR (red) and WSMR (blue) as well as all three events in the synthetic data set (NNSS in green). Solid bars denote the number of realizations in which each event cluster was incorrectly defined (corrupted event clusters), and the patterned bars denote the number of realizations in which noise detections were incorrectly included in the event clusters (false associations).

Figure 11. Backazimuth projections for the events from Fig. 9 with biases removed. All projections intersect at the true source location in each case.

Figure 12. Cluster labels identified for 100 samples from adding different levels of perturbations to the detection backazimuths using the same representation as in Fig. 10.

It should be noted that the eigenray analysis used to estimate propagation path characteristics produced a detection set with a mean and maximum azimuth deviation of 3.4° and 8.5°, respectively, yet the clustering algorithm is able to easily identify the three events. This is likely due to the nature of backazimuth biases introduced due to propagation effects. For a source observed perpendicular to strong cross winds, the apparent source direction is shifted in the direction of the winds regardless of the direction of propagation. That is, summarizing the discussion by Blom & Waxler (2017), for a strong westward stratospheric wind, observations to both the north and south are expected to observe an apparent source shifted to the west of the true source location. This implies that backazimuth
4.2 Sensitivity to errors in detection arrival time

In addition to the errors in the backazimuth of detections, the long duration and emergent nature of many infrasonic signatures can make identification of a precise arrival time a challenge in some cases. In addition, detections can be defined by the time of the signal onset or by the time of the peak coherence or Fisher statistic value. Therefore, it is useful to quantify the sensitivity of the association methods developed here to errors in the detection time. Following an analysis similar to that utilized to investigate the sensitivity to errors in the backazimuth, perturbations to the arrival times of the detections have been generated using a uniform distribution on an interval $\pm \delta \tau$ to produce 100 realizations of the detection set with the non-noise detection arrival times shifted by some amount. For this analysis, the original set of detections including propagation related backazimuth deviations has been used and the same set of noise detections as listed in Table 4 and utilized in the backazimuth sensitivity analysis are included in the data set.

The resulting labels and clustering summary are shown in Fig. 13 for arrival time perturbations up to $\pm 120$ s and are unsurprising given the discussion of other results. Even for unrealistically large uncertainties of 90 or 120 s, the association algorithm is able to consistently identify every case including only two events and in more than 90 per cent of the realizations using all three sources. This is due to the large uncertainties included in the celerity model as mentioned above and demonstrate just how important of a role accurate backazimuth estimation has in infrasonic signal analysis. While reasonable uncertainties in arrival time are overwhelmed by uncertainties in propagation time, the nature of acoustic propagation, even in cases with relatively strong cross winds or other effects that influence backazimuth, makes the observed backazimuth significantly more important in identifying events and localizing and characterizing sources.

5 CONCLUSIONS

A pair-based, Bayesian association algorithm has been developed that leverages the joint-likelihood between pairs of infrasonic detections to identify events within a detection list from a network of infrasound stations. The method uses the negative log-joint-likelihood to define a non-Euclidean distance matrix for event identification. Infrasonic propagation models for backazimuth projection and propagation time are incorporated into the likelihood function definition to ensure physically realistic results are obtained in a computationally efficient manner. A weighted, agglomerative clustering method is applied to the distance matrix defined by the negative log-joint-likelihood of detection pairs and events are identified by a user defined distance threshold and minimum cluster size criterion. An automated trimming algorithm has been developed to identify and correct poorly shaped clusters for which strong association between subsets of the cluster members and non-member detections are identified. This initial implementation of the method utilizes likelihoods defined so that only spatial and temporal coincidence are used to associate detections; however, additional measures including estimated source strength could be included in future work.

A synthetic data set has been constructed using a network of four infrasound arrays in the western United States and three source locations. Two sources, NNSS and UTTR, are spatially similar and introduce a challenge in resolving the two separate events in the analysis. A third source, WSMR, is located near the edge of the network and has poor azimuthal coverage in order to determine what influence limited network coverage has on the association methodology. In addition to infrasonic signals produced by the three sources, noise detections have been added to the data set to evaluate the method’s performance in more realistic scenarios. Events are correctly identified in approximately 85–95 per cent of cases depending on the source and network geometry. In both
cases, noise detections are incorrectly included in the events in 15–20 per cent of cases, which could likely be reduced by replacing the automated cluster trimming algorithm with an analyst review once the distance matrix has been computed and clustering has been performed.

In addition to studying performance of the method for various noise realizations, sensitivity studies have been conducted for variations in detection backazimuths and arrival times in order to determine how propagation- or detection induced errors will degrade association analysis results. Backazimuth errors up to ±8° are found to produce negligible errors in cluster identification for a pair of spatially separated sources. For all three events, spatially similar events become mixed due to backazimuths shifting between similar source azimuths. For backazimuth errors up to ±8°, spatially similar events become mixed or otherwise incorrectly identified in 15–20 per cent of cases. Larger backazimuth deviations produce a significant increase in mixing and incorrect event identification for spatially similar events due to the lack of azimuthal resolution for some of the network arrays and the limited azimuth uncertainty used in defining the spatial integration region used to compute the joint-likelihood. Source time errors have little influence on the association results due to the large uncertainty in propagation time for the celerity model utilized in the likelihood definition. Even for overly large arrival time errors of 90–120 s, spatially separate events are found to be consistently identified correctly and spatially similar events are mixed or incorrectly identified in less than 10 per cent of cases.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of Ms Leslie Casey and the National Nuclear Security Administration Office of Nonproliferation and Treaty Verification Research and Development for funding the research discussed here. Los Alamos National Laboratory completed this work under the auspices of the U.S. Department of Energy.

REFERENCES


