

1 Automatic Slowness Vector Measurements of Seismic Arrivals
2 with Uncertainty Estimates using Bootstrap Sampling, Array
3 Methods and Unsupervised Learning

4 J. Ward¹ M. Thorne² A. Nowacki¹ S. Rost¹

5 ¹ *School of Earth and Environment, University of Leeds*

6 ² *Department of Geology and Geophysics, University of Utah*

5 SUMMARY

6

7 Horizontal slowness vector measurements using array techniques have been used to analyse
8 many Earth phenomena from lower mantle heterogeneity to meteorological event location.
9 While providing observations essential for studying much of the Earth, slowness vector analy-
10 sis is limited by the necessary and subjective visual inspection of observations. Furthermore, it
11 is challenging to determine the uncertainties caused by limitations of array processing such as
12 array geometry, local structure, noise and their effect on slowness vector measurements. To ad-
13 dress these issues, we present a method to automatically identify seismic arrivals and measure
14 their slowness vector properties with uncertainty bounds. We do this by bootstrap sampling
15 waveforms, therefore also creating random sub arrays, then use linear beamforming to mea-
16 sure the coherent power at a range of slowness vectors. For each bootstrap sample, we take
17 the top N peaks from each power distribution as the slowness vectors of possible arrivals. The
18 slowness vectors of all bootstrap samples are gathered and the clustering algorithm DBSCAN
19 (Density-Based Spatial Clustering of Applications with Noise) is used to identify arrivals as
20 clusters of slowness vectors. The mean of each cluster gives the slowness vector measurement
21 for that arrival and the distribution of slowness vectors in each cluster gives the uncertainty
22 estimate. We tuned the parameters of DBSCAN using a dataset of 2489 SKS and SKKS ob-

23 servations at a range of frequency bands from 0.1 Hz to 1 Hz. We then present examples at
24 higher frequencies (0.5 to 2.0 Hz) than the example dataset, identifying PKP precursors, and
25 lower frequency by identifying multipathing in surface waves (0.04 to 0.06 Hz). While we use
26 a linear beamforming process, this method can be implemented with any beamforming pro-
27 cess such as cross correlation beamforming or phase weighted stacking. This method allows
28 for much larger datasets to be analysed without visual inspection of data. Phenomena such as
29 multipathing, reflections or scattering can be identified automatically in body or surface waves
30 and their properties analysed with uncertainties.

31 **Key words:** Body waves, Surface waves and free oscillations, Structure of the Earth

32 1 INTRODUCTION

33 Seismic array techniques which measure the full horizontal slowness vector (backazimuth and
34 inclination) of seismic arrivals have been used to investigate Earth structure for decades. These
35 analyses have been applied to a wide variety of seismic arrivals and problems such as by using
36 long period surface waves to identify upper mantle and surface heterogeneity (Ji et al., 2005;
37 Maupin, 2011; Xia et al., 2018), short period S-waves to analyse lower mantle structure (Cottaar
38 & Romanowicz, 2012; Schumacher & Thomas, 2016; Stockmann et al., 2019; Ward et al., 2020),
39 high-frequency P-waves to study scatterers in the mid and lower mantle (Niu & Kawakatsu, 1997;
40 Thomas et al., 2002; Cao & Romanowicz, 2007; Frost et al., 2013; Bentham & Rost, 2014; Yang
41 & He, 2015; Ritsema et al., 2020), event detection and spatial location (Chevrot et al., 2007;
42 Landès et al., 2010; Liu et al., 2016), ambient noise (Behr et al., 2013; Roux & Ben-Zion, 2017),
43 nuclear event detection (Bowers & Selby, 2009; Gibbons & Ringdal, 2011) and meteorological
44 event spatial location (Gerstoft et al., 2006, 2008).

45 Past studies which analysed slowness vector properties using array methods (for a review see:
46 Rost & Thomas, 2002, 2009) were limited in terms of number of observations due to the usual
47 requirement to visually inspect each observation to determine an arrivals slowness vector proper-
48 ties or if it is too noisy to use. In addition, several studies have discussed the limitations of using
49 beamforming or f–k methodology to identify phases and estimate their slowness vector properties

Automatic Slowness Vector Measurements of Seismic Arrivals with Uncertainty Estimates using Unsupervised Learning

50 (Berteussen, 1976; Gibbons et al., 2008; Selby, 2011) and methods have been developed to correct slowness vector measurements for Earth structure when locating events (Bondár et al., 1999; 51 Koch & Kradolfer, 1999; Schweitzer, 2001; Gibbons et al., 2011). To clarify what limitations the 52 uncertainty estimate is accounting for, we first discuss the assumptions and limitations of making 53 one slowness vector measurement.

56 Each slowness vector in the beamforming grid search assumes the wave moves over the array 57 with a constant horizontal slowness and arrives at the stations with a backazimuth equal to that 58 along the great circle path from the relocated event location (for details see: Ward et al., 2020). 59 The beamforming process does not account for limitations in heterogeneous station distribution, 60 which can lead to heterogeneous sampling of the wavefield, and interference from noise may 61 contribute to errors in slowness vector measurement. The waveforms of the arrivals are assumed 62 to be coherent across over the array. Waveform incoherence of the signal across the array may 63 result in deviations from the slowness vector prediction (Gibbons et al., 2008). Source complexity 64 could lead to unusual waveforms recorded at the array, but should not affect the slowness vector 65 measurement as source complexity should introduce consistent waveform complexity across the 66 array. Local structure may deform the wavefield as it moves across the array such that the arrival 67 times at the stations will deviate from the prediction (Gibbons et al., 2018). This may lead to 68 slowness vector deviations depending on geometry of stations distribution and local velocity and 69 topography structure.

70 The predicted backazimuth of the arrival is assumed along the great circle path between the 71 event and the mean station location assuming a spherical Earth. The predicted horizontal slowness 72 of the arrival is taken from ray tracing through a 1-D velocity model in a spherical Earth. Any 73 structures local to the array or deeper with properties which differ from the 1-D velocity model 74 may result in deviations from this prediction. It is difficult if not impossible to separate out these 75 different contributions using just a single array measurement, let alone determine their relative 76 contributions.

77 Automating the identification of arrivals and measuring their slowness vector properties would 78 remove the time consuming and subjective process of visually inspecting each observation and

79 could allow for larger data sets to be analysed. Estimating the uncertainty of these measurements
80 allow for better interpretation of the observations, and the ability to rigorously accept or reject
81 scientific hypotheses on Earth structure or its processes.

82 Previous efforts have been made in automating standard seismic processing techniques such
83 as shear wave splitting (Teanby et al., 2004) and $H - \kappa$ stacking (Ogden et al., 2019). Methods
84 also exist to estimate uncertainties in the beamforming methodology (Lin & Roecker, 1996; Bear
85 & Pavlis, 1997; Ritsema et al., 2020) and to improve the detection of one or multiple arrivals (Gal
86 et al., 2014, 2016; Schmidt, 1986). The method we propose differs from these by automatically
87 identifying the number of arrivals with their slowness vector properties and uncertainties. To our
88 knowledge, no method has been proposed that does all of these at once. The method we present
89 later uses a linear relative beamforming process; however, this method can be applied with other
90 techniques such as phase weighted stacking (Schimmel & Paulssen, 1997) or cross correlation
91 beamforming (Ruigrok et al., 2017).

92 Machine learning methodologies are becoming more prevalent in the geosciences (for a review
93 see: Bower et al., 2013) and seismology (for a review see: Kong et al., 2019) with methods used
94 to automate data selection (e.g. Valentine & Woodhouse, 2010; Thorne et al., 2020) and extracting
95 properties from data by mapping seismograms to lower dimensional space using autoencoders
96 (Valentine & Trampert, 2012) or sequence seismograms and identify features such as the presence
97 of seismic scatterers (Kim et al., 2020). Here we use an unsupervised learning algorithm as part of
98 our automation technique.

99 In the approach we present in this paper, we create subsets of waveforms using bootstrap
100 sampling (Efron, 1992). For each sample, beamforming (Rost & Thomas, 2002) corrected for a
101 curved wavefront (Ward et al., 2020) is used to search over a range of slowness vectors and re-
102 cover the slowness vectors of potential seismic arrivals. The slowness vector measurements of all
103 the individual bootstrap samples are collected and we use the DBSCAN (Density-Based Spatial
104 Clustering of Applications with Noise) algorithm (Ester et al., 1996) to identify clusters of slow-
105 ness vectors as seismic arrivals. DBSCAN is an unsupervised learning algorithm which uses the
106 density of points to classify them as part of a cluster or as noise. For further details, see Section 2.

Automatic Slowness Vector Measurements of Seismic Arrivals with Uncertainty Estimates using Unsupervised Learning

By bootstrapping the traces, and therefore creating random subsets of the stations in the array, the scatter of the measurements in each cluster can give an estimate of the combination of some of the previously mentioned uncertainties. The uncertainty estimate will account for the following phenomena which cause different subsets of stations to have different slowness vector measurements:

- signal aberration where the arrival time of the wave at stations deviates from the prediction due to local array structure;
- incoherent or coherent noise;
- the horizontal slowness of the wave changing as it moves over the array, due to the size of the array, or unaccounted for velocity variations within the array;
- heterogeneous distribution of the stations causing heterogeneous sampling of the wavefield;
- slowness resolution limitations of the array aperture; and
- wavelet shape changing over the array.

All of these can relate to local structure or effects within the array and the uncertainty estimate describes the combination of all effects on the wavefield. If a measured slowness vector deviates from the 1-D Earth model prediction and is not within the uncertainty estimate, then the cause of this deviation must be external to the array and local structure. Determining the cause of these deviations to structures such as a dipping Moho, or deeper structure requires additional information and might be resolvable through e.g. forward modelling. We do not try to measure the uncertainties of that aspect, only those listed above.

We tune the parameters of the DBSCAN algorithm on a visually inspected dataset where each observation is labeled as having either 0, 1, or 2 arrivals. More arrivals are possible, but in this dataset the maximum number confidently observed is 2. In this dataset, observations with more than one arrival are hypothesised to be caused by multipathing, one of many phenomena which can cause multiple arrivals. Multipathing occurs when the wavefront is incident of a sufficiently large velocity gradient causing different parts of the wavefield to move at different velocities, diffract and refract. Multipathing results in 2 arrivals arriving at the station at different times and different slowness vector properties. The predictions made by the method are compared to the labels given

136 from visual inspection to find the best parameters for the DBSCAN algorithm. Following this, we
137 show the effectiveness of this automated method on finding the slowness properties of short-period
138 PKP scattering and long-period surface wave arrivals. Guidance on using the method is given in
139 Section 5. We find the parameters work well for our example applications with a minor change
140 needed for the surface wave example. Tuning the algorithm can be done for specific applications.

141 **2 METHOD OVERVIEW**

142 This section outlines the method to automatically measure the slowness vector properties with
143 uncertainty estimates. The process can be roughly broken down into the following steps with more
144 detail given below.

145 (i) Create a number of bootstrap sub-samples (1000 here) through random sampling with re-
146 placement of a set of waveforms recorded at the seismic array in question.

147 (ii) For each bootstrap sample, use beamforming (Rost & Thomas, 2002) correcting for a
148 curved waveform (Ward et al., 2020) to search over a grid of slowness vectors and find how the
149 power of coherent energy varies with backazimuth and horizontal slowness. Therefore, each boot-
150 strap sample will have its own grid of power values.

151 (iii) Calculate a noise estimate for the bootstrap sample by shifting each trace in the bootstrap
152 sample with a randomly generated time. These scrambled traces are then stacked and the power
153 of the beam is measured. This is repeated 1000 times and the mean power is taken as the noise
154 estimate.

155 (iv) Set all power values in the slowness grid below the noise estimate to zero.

156 (v) From the resultant power distribution, take up to X peaks (in this study we take up to 3
157 peaks), which describe the slowness vectors of possible arrivals.

158 (vi) Gather the locations for these peaks of all the bootstrap samples.

159 (vii) Use DBSCAN, a density-based clustering algorithm, to identify the arrivals and measure
160 their slowness properties with uncertainties.

161 2.1 Bootstrapping and peak recovery

162 One advantage of the bootstrap sampling process is that bootstrap samples of the stations in the
163 array are used. Beamforming subsets of the array leads to different peak power in the beams which
164 leads to variations in the recovered slowness vectors for each arrival. When all of the slowness
165 vectors are taken into account, using all of the bootstrap sampled arrays, we obtain uncertainty
166 estimates in the slowness vector. These uncertainty estimates will include the effect that array
167 geometry and local structure has on the slowness vector measurements. For each bootstrap sample,
168 we use a relative beamforming method where the traces are aligned on a target slowness before
169 searching over the slowness vectors. After the beamforming, we calculate a noise estimate using
170 the traces in the bootstrap sample with a similar method to Korenaga (2013). The traces are aligned
171 using the slowness vector with the highest power. Then, they are randomly shifted in time, stacked
172 and the power of the stack calculated. This is repeated 1000 times and the mean of all power
173 estimates is used for the noise power estimate. All power values in the beamforming plot (Fig
174 1) below three times this noise estimate are set to zero. Multiplying the estimate by three was
175 determined by exploratory analysis and found to give the most satisfactory result. This can be
176 changed depending on the application. To remove local power maxima, the power distribution is
177 smoothed using a 2-D Gaussian filter. The 2-D Gaussian is formed by the product of two 1-D
178 Gaussians. The standard deviation of the 1-D gaussians is equal to the grid spacing (0.05 s°),
179 therefore will have a full width at half maximum of 0.12 s° .

181 The 2-D Gaussian acts as a point spread function and is convolved with the power plot to
182 smooth it and remove local maxima. After this, the top X peaks are taken from the power distri-
183 bution. The peaks are found with a maximum neighbourhood filter which identifies points with
184 higher power values than those in the surrounding neighbourhood. Fig 1 shows how the peaks are
185 found for each bootstrap sample.

186 2.2 Identifying arrivals with cluster analysis

187 The peaks recovered for each bootstrap sample are then collected and the clustering algorithm DB-
188 SCAN (Ester et al., 1996) is used to find clusters. DBSCAN is an unsupervised learning algorithm

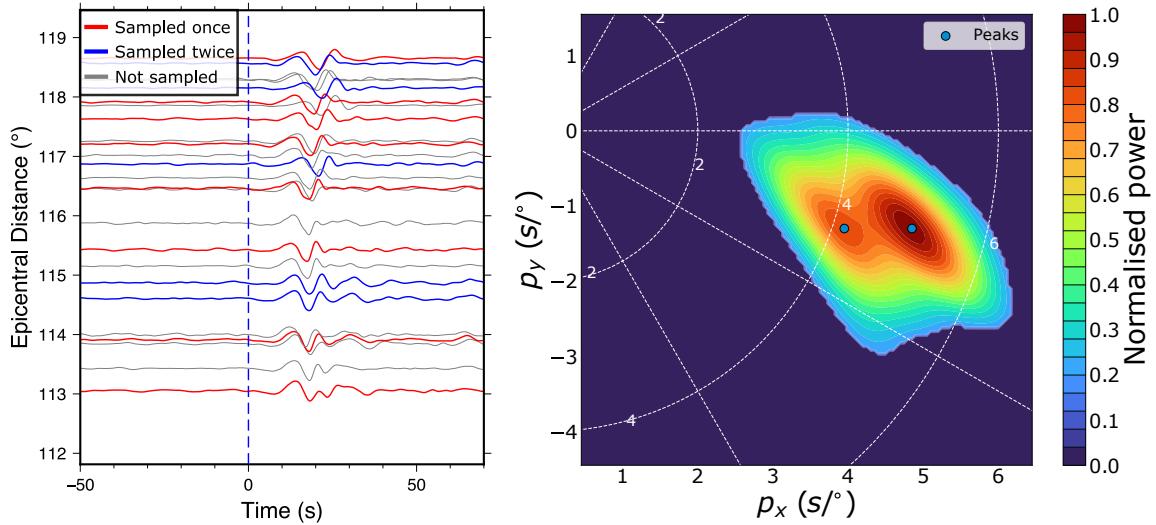


Figure 1. Example of recovery of peaks from a bootstrap sample of traces. The left figure shows a record section of data from the 05, April 1999 event recorded at the Kaapvaal array in Southern Africa (event metadata in the supplementary material). The traces are coloured by the number of times they have been sampled. The data had the instrument response removed and are filtered between 0.10 and 0.40 Hz before beamforming. The right figure shows the power distribution at each slowness vector with powers lower than the noise estimate set to zero and the 2-D Gaussian smoothing filter applied. Here each point on the grid represents a slowness vector described with their x (p_x) and y (p_y) components. In this example, two peaks have been recovered.

which uses the density of points to identify clusters and noise. The algorithm takes a radius ϵ and a minimum number of points (MinPts) to define a minimum density for points to be a cluster. Here, we define MinPts as a fraction of the number of bootstrap samples. DBSCAN sorts the data into three categories as visualised in Fig 2.

- (i) Core point: A point with at least MinPts points within its neighbourhood (i.e. within radius ϵ).
- (ii) Boundary point: A point within the neighbourhood of a core point, but without MinPts points in its own neighbourhood.
- (iii) Noise: Points that are not within ϵ of a core point and does not have MinPts points within its neighbourhood.

The DBSCAN algorithm begins at a random point and measures its density by the number of points within the radius ϵ (Fig 2). If the density is lower than the threshold defined by ϵ and

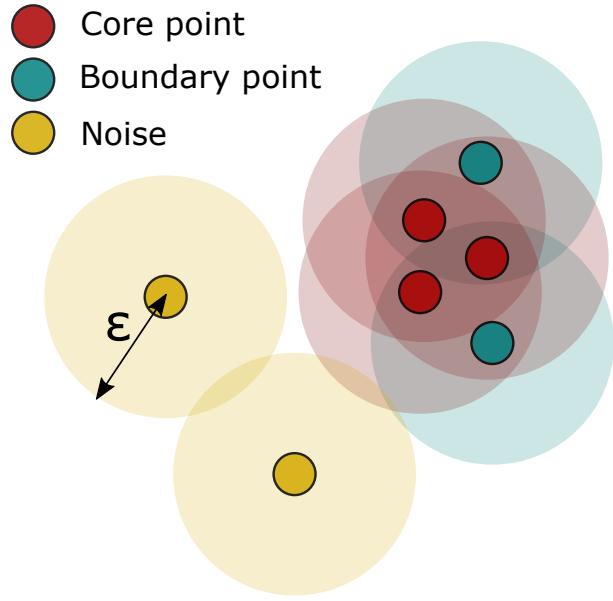


Figure 2. Cartoon illustrating what classifies as a core point, boundary point or noise. The neighbourhoods of the points are shown as a lighter colour of the point itself. The minimum number of points needed for a core point is 4 in this example. The red points all have at least 4 points in their neighbourhood, so are defined as core points. The blue points are within the neighbourhood of the core (red) points, but do not have 4 points in their own neighbourhood and are classified as boundary points. The yellow points are classified as noise because they are not in the neighbourhood of a core point and do not have 4 points within their own neighbourhood.

201 MinPts, the point is classified as noise (yellow points in Fig 2) and the algorithm moves on to
 202 another random point. If the density is higher than the defined threshold, the point is classified as
 203 a core point and cluster formation begins (red points in Fig 2). Points within ϵ of the core points
 204 then have the number of points in their neighbourhood measured. Those which do not have MinPts
 205 points within their neighbourhood are boundary points and are still part of the cluster (blue points
 206 in Fig 2). The points which do have MinPts points in their neighbourhood are classified as core
 207 points and added to the cluster. The points within ϵ of these new core points are also searched and
 208 the cluster expands until it finds no new core points to add to the cluster. Once no new core points
 209 can be added, an unexamined point is chosen at random and the process begins again. This process
 210 continues until all points have been examined. In this manner, DBSCAN can separate high density
 211 clusters from low density noise. Fig 3 shows the result of DBSCAN applied to the peaks recovered
 212 after the bootstrapping process.

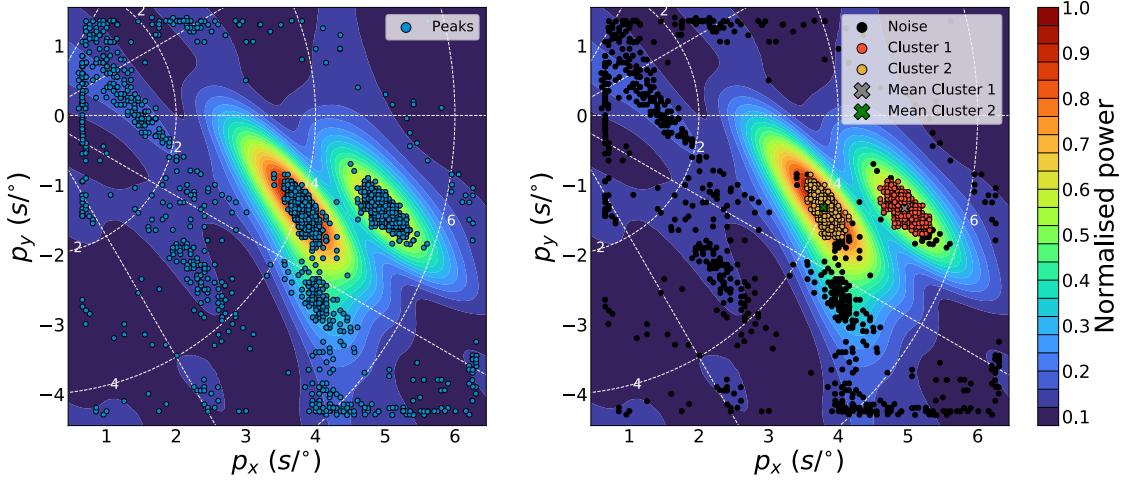


Figure 3. Cluster retrieval from points recovered through bootstrap sampling the traces (Fig 1). The left figure shows all the power peaks (blue dots) recovered using data from the 05 April 1999 event. The right image shows the clusters found by the DBSCAN algorithm (Ester et al., 1996) where MinPts is 0.25 and ϵ is 0.2 s/° . The red and yellow points are classified as clusters 1 and 2 respectively and the black points are noise. The background power distribution is the mean of all the power distributions found from bootstrap sampling.

DBSCAN has advantages over other clustering algorithms such as k -means (MacQueen et al., 1967) for this application such as:

- (i) It does not take the number of clusters as input so visual inspection before the clustering is not required.
- (ii) Not all points need to be part of a cluster allowing for noise.
- (iii) If clusters are not well separated or the data is noisy, clusters of non-hyperspherical shape can still be recovered unlike k -means (Ertöz et al., 2003; Celebi et al., 2013).

There are also disadvantages to DBSCAN:

- (i) If the range and data is not well understood, choosing the parameters can be challenging.
- (ii) Clustering data with large variations in density is challenging because there may be no combination of ϵ and MinPts which will find all of the clusters.
- (iii) Clusters separated by a distance smaller than ϵ will be combined into one cluster.

We tested other density-based clustering algorithms such as HDBSCAN (Campello et al.,

2013, 2015) and OPTICS (Ankerst et al., 1999) but found that both techniques have issues for this application. HDBSCAN (Hierarchical DBSCAN) searches over a range of ϵ values and measures over what length scales a cluster “persists” while containing a minimum number of points to form a cluster. Using how long each cluster survives and how many points it contains at each ϵ , clusters are extracted with the excess of mass algorithm (EOM) (McInnes & Healy, 2017). HDBSCAN will preferentially return a large, single cluster because one large cluster will usually contain more “mass” (for a detailed explanation, see McInnes & Healy, 2017). To avoid one large cluster being returned when multiple clusters exist, HDBSCAN by default will not return a single cluster as an output. If this default is kept, instances with one arrival (cluster) will be misidentified. Changing the default and allowing HDBSCAN to return one cluster will mean phenomena causing multiple arrivals (such as multipathing) may not be identified as EOM will preferentially return a single cluster.

OPTICS (Ordering Points To Identify the Clustering Structure) (Ankerst et al., 1999) is another density-based algorithm which specialises in identifying clusters of varying density. OPTICS orders the points to represent the clustering structure. From this, clusters can be extracted. When using OPTICS, we found the size of the clusters retrieved was too inconsistent to estimate the uncertainties of slowness vector properties. Because of these considerations, we decide to use DBSCAN instead of OPTICS or HDBSCAN.

2.3 Slowness Vector Uncertainty Estimates

We estimate the uncertainty with the standard deviation of backazimuths and horizontal slownesses in each cluster and also use the area of error ellipse of the clusters as a relative measure of uncertainty of each observation. The error ellipses are found by calculating the eigenvectors and eigenvalues of the covariance matrix for each cluster. These eigenvectors and eigenvalues give the directions and magnitudes of the maximum variances in the cluster which is used to determine the width, length and orientation of the ellipse. Fig 4 shows clusters plotted with their error ellipses for 1, 2 and 3 standard deviations. We would like to highlight the importance of the slowness grid

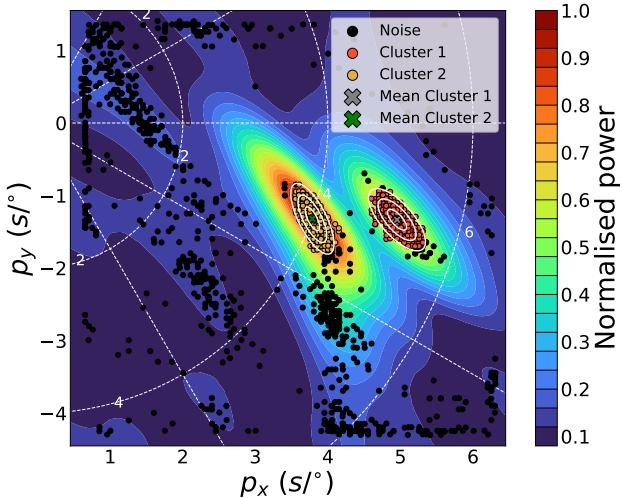


Figure 4. Example of error ellipses for 1,2 and 3 standard deviations. The data are the same as used in Figs 1 and 3. The background power plot is the mean of the power plots searching over a range of slowness vectors from each bootstrap sample.

254 dimensions; if the slowness grid is too small, the arrivals may be truncated at the edge leading to
 255 a smaller cluster and underestimate the uncertainty.

256 3 PARAMETER TUNING

257 To find the best parameters to use with the DBSCAN algorithm (ϵ and MinPts), we compare
 258 the number of arrivals predicted by the algorithm to the number of arrivals identified from visual
 259 inspection. We use the same dataset as Ward et al. (2020) which used SKS and SKKS data recorded
 260 at the Kaapvaal array in southern Africa. Ward et al. (2020) make observations at a range of
 261 frequency bands (Table 1) using the whole Kaapvaal array and several sub-arrays. The traces are
 262 first aligned on the predicted slowness of SKS or SKKS depending on the arrival of interest. The
 263 beamforming is conducted in a time window that is 20s before and 40s after the predicted arrival.

264 The dataset provides a good test for the algorithm since it has clear single arrivals, multipathed
 265 arrivals (2 arrivals) and observations that are too noisy to identify any arrivals (0 arrivals). Each
 266 observation is labeled from visual inspection of the distribution and density of the points collected
 267 from all the bootstrap samples and the mean power distribution of all the bootstrap samples. If
 268 the algorithm predicts a higher number of arrivals than the human given labels, we assume here

Table 1. The number of labels in each frequency band. Labels indicate the number of arrivals in that observation and 1-2 could be either 1 or 2. In total, there are 2628 labels with 2489 used in the tuning.

Frequency (Hz)	Number of Arrivals			
	1	2	1-2	0
0.07 - 0.28	403	18	10	7
0.10 - 0.40	378	21	20	19
0.13 - 0.52	326	33	25	54
0.15 - 0.60	308	28	23	73
0.18 - 0.72	280	27	27	104
0.20 - 0.80	253	35	28	122
Total	1948	162	133	379

269 the algorithm has identified noise as arrivals. If the algorithm predicts a lower number of arrivals,
 270 the density threshold is too high for arrivals to be identified. Due to the subjective nature of the
 271 labelling this may not always be the case, but for the tuning process we assume the human labels
 272 are a ground truth. Observations where it was not clear whether there is one or two arrivals are
 273 labeled as “1-2 arrivals” and excluded from this tuning process.

275 We searched over a range of ϵ and MinPts values and predict the number of arrivals in each
 276 observation. This is compared to the human labels in Table 1 and an accuracy score is calculated.
 277 The accuracy score is defined as the number of instances where the method correctly predicts
 278 the number of arrivals relative to the total number of instances ($\frac{\text{No. correct predictions}}{\text{Total instances}}$). Values of ϵ
 279 range from 0.05 to 1.0 s/ $^{\circ}$ and MinPts is given as a fraction of the bootstrap samples (1000 here)
 280 and varies from 0.05 to 1.0. Fig 5 shows how the accuracy varies in the parameter space. The
 281 grid search shows the sensitivity of our method to the DBSCAN parameters chosen. With some
 282 parameters, the accuracy can exceed 90 % while with others it can be less than 20%. The method
 283 performs the worst with small ϵ and high MinPts meaning the minimum density criteria will be
 284 very high and very few arrivals will be found.

285 We test how well the algorithm generalises using cross validation. Cross validation involves

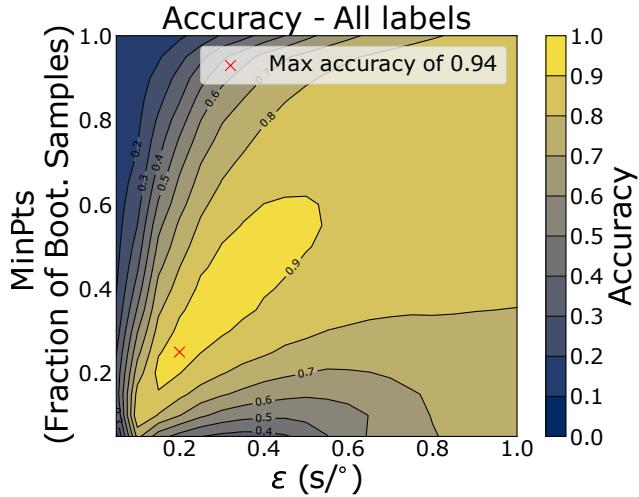


Figure 5. Grid search of DBSCAN parameters ϵ and MinPts (given as a fraction of bootstrap samples). For each combination, the number of arrivals in each observation are predicted, compared to the true labels (Table 1) and the accuracy calculated. The location of the highest accuracy value is plotted as a red cross where $\epsilon = 0.20 \text{ s/}^\circ$ and MinPts = 0.25.

286 splitting the dataset into N representative subsets (5 here). One of the subsets is removed and the
 287 grid search is conducted on the remaining $N - 1$ subsets and the best set of parameters recorded.
 288 The removed subset acts as a validation set. Then we take these best parameters and make predic-
 289 tions on the validation set. The accuracy of the predictions for the validation subset is measured
 290 and gives an indication of how well the algorithm generalises. The process is repeated by sequen-
 291 tially removing one subset and tuning the parameters on the remaining $N - 1$ subsets. After the
 292 cross validation process, there are N estimates indicating how well the algorithm performs on
 293 unseen data. Here we split the data into 5 subsets because of the low number of multipathed (2
 294 arrivals) and 0 arrivals samples. Cross validation and measuring the accuracy gave a mean accu-
 295 racy of 0.939 with a standard deviation of 0.0090. In all the cross validation samples, the best
 296 parameters were $\epsilon = 0.20 \text{ s/}^\circ$ and MinPts = 0.25.

297 As there are many more instances of observations with one arrival, we also analyse each of the
 298 target labels (0,1 or 2 arrivals) individually using the precision, recall and F1 measures (defined
 299 below). These measures all depend on the number of true positive (TP), true negative (TN), false
 300 positive (FP) and false negative (FN) instances. These are best understood with an example. If the
 301 target label is “2”, true positives are instances where the algorithm correctly identifies 2 arrivals

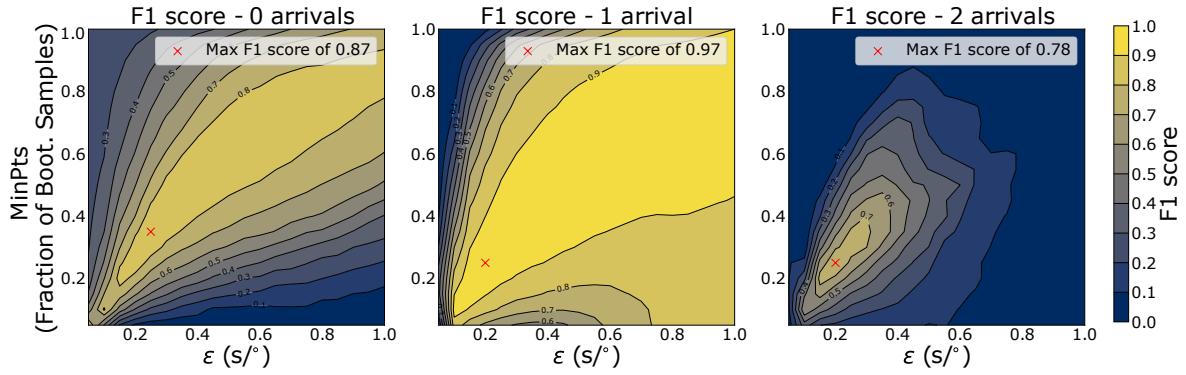


Figure 6. F_1 scores for combinations of DBSCAN parameters ϵ and MinPts where each plot represents a different target labels of 0 arrivals (left) one arrival (centre) and two arrivals (right). The location of the highest F_1 score is plotted as a red cross, which has parameters of $\epsilon = 0.20$ s/ $^\circ$ and MinPts = 0.25 for 1 and 2 arrivals and $\epsilon = 0.25$ s/ $^\circ$ and MinPts = 0.35.

in an observation. True negatives are instances correctly identified as not having 2 arrivals (1 or 0 arrivals). False positives are those incorrectly identified as having 2 arrivals. False negatives are instances where 2 arrivals have not been identified when they should have been.

From these measures, the precision is defined by $P = \frac{TP}{TP+FP}$. This is essentially the proportion of the target labels which have been correctly identified. The recall, $R = \frac{TP}{TP+FN}$, is a measure of how many of the target labels has been recovered by the algorithm. The F_1 score is the harmonic mean of the precision and recall and can be described as $F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}}$. The F_1 score is only large if both the recall and precision are high. We only present the F_1 score as it shows which parameters have both high precision and recall. Fig 6 shows how the F_1 score varies with different parameter combinations for each target label.

Figures 5 and 6 show that the method is capable of greater than 90% agreement with the observations of a human. This is mainly from observations with one clear arrival, which makes up the majority of the observations. The algorithm also performs well with more complex observations of multipathing with a F_1 score of over 0.75. This method is quite insensitive to noise as it does not regularly incorrectly identify noisy observations as shown by a F_1 score of over 0.85 for observations with 0 arrivals. As with the accuracy, we use cross validation to see how well the parameters generalise with new data. Table 2 shows the mean F_1 scores for the individual labels.

Table 2. Table of the cross-validation result for each of the labels (0,1 or 2 arrivals) where the F_1 score is the measure of success. Notice the standard deviation is an order of magnitude higher for labels 0 and 2, most likely because of the significantly fewer instances of those labels in the subsets created during cross-validation.

No. Arrivals	Mean F_1 score	Standard Deviation	Best Parameters
0	0.86	0.030	$\epsilon = 0.35 \text{ s}^\circ$ MinPts = 0.25
1	0.97	0.0063	$\epsilon = 0.20 \text{ s}^\circ$ MinPts = 0.25
2	0.78	0.035	$\epsilon = 0.20 \text{ s}^\circ$ MinPts = 0.25

320 As in Figure 5, there are DBSCAN parameters which perform very poorly showing the importance
 321 of the parameters used.

322 The cross validation analysis of all the labels and F_1 score on the individual labels show the
 323 parameters $\epsilon = 0.20 \text{ s}^\circ$ and MinPts = 0.25 are consistently found to be the best. Inferring how
 324 well the parameters generalise from this analysis is limited because of the low number of cross
 325 validation samples (5 here). The low sample number was necessary because of the small number
 326 of observations with 2 and 0 arrivals. Despite this, the mean values obtained for the accuracy score
 327 and F_1 scores from the cross validation are very similar to that obtained by tuning with all the
 328 data (Figs 5 and 6). The standard deviations from the cross validation are low suggesting similar
 329 performance on similar datasets.

330 Due to the subjective nature of labelling each observation with the number of arrivals, some
 331 difference between the method's prediction and the human labels is acceptable. To analyse how
 332 reasonable the predictions are when the technique disagrees with the human labels, we create a
 333 confusion matrix using the predictions with parameters of $\epsilon = 0.20 \text{ s}^\circ$ and MinPts = 0.25 (Fig 7).
 334 In the confusion matrix, each row represents a true label (number of arrivals in this case) and each
 335 column the predicted arrivals. The values at each point in the matrix indicates how many times
 336 that true label is identified as the corresponding predicted labels. For example, for all instances

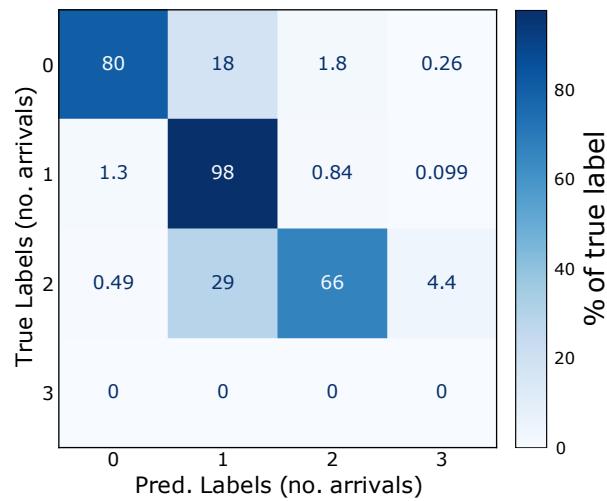


Figure 7. Confusion matrix for predictions made with $\epsilon = 0.20 \text{ s}^\circ$ and $\text{MinPts} = 0.25$. Each row represents a true label (number of arrivals) and each column the predicted arrivals. The values on the diagonal of the matrix show the percentage of correct predictions for the true label.

with the true label of 1 arrival, the confusion matrix will show how many are correctly classified as having one arrival and how many are incorrectly identified with 0, 2 or 3 arrivals. We normalise the values along each row of the confusion matrix so for each true label, the columns show the proportion of the predictions given to that label. For example, for the instances with a true label of ‘0 arrivals’, 80 % of the predictions are correctly identified as having 0 arrivals, 18 % are identified as having 1 arrival and so on.

The confusion matrix shows that when the method prediction differs from the human labels, the predictions it makes are not radically unreasonable. It is worth remembering the labeling process is quite subjective and just because the algorithm predicts a different number of arrivals to that labeled by a human, does not mean it is wrong. It is possible that some of the human labels with two arrivals only have one arrival or some have three arrivals. Equally, it is possible some instances labeled with no arrivals do have one arrival but a human could not confidently identify it above the noise. Fig 7 shows the algorithm makes reasonable predictions in the vast majority of the cases for this data set using the parameters found from the tuning process and cross validation. Analysis of the uncertainty estimates show the slowness vector measurements have small variation with the mean standard deviation for backazimuth measurements of 1.2° and horizontal slowness of 0.14 s° . The mean area bounded by the 95% confidence ellipse is $0.14 \text{ s}^2/\text{^\circ}$.

355 Analysis of the confusion matrix in addition to the findings from the cross validation process
356 shows the parameters $\epsilon = 0.20$ s $^{\circ}$ and MinPts = 0.25 will give reasonable results that will generalise
357 well. We use this parameters in other applications with a minor change for applications to surface
358 waves (Section 4).

359 4 APPLICATIONS TO PKP SCATTERING AND RAYLEIGH WAVE MULTIPATHING

360 This section provides two example applications of this method to study Earth structure. First, we
361 show an example identifying a PKP precursor in the high frequency teleseismic wavefield (0.5 to
362 2 Hz). Coherent precursors are indicative of scattering caused by small scale structures and our
363 method can constrain uncertainties on their location. Then, we show an example of low frequency
364 (0.04 to 0.06 Hz) Rayleigh wave multipathing. Using our method to identify Rayleigh wave mul-
365 tipathing, we can interpret possible causes of multipathing and provide uncertainties for phase ve-
366 locity measurements. All measurements of backazimuth and horizontal slowness are shown with
367 one standard deviation describing the uncertainties.

368 4.1 PKP precursors

369 Analysing the slowness vectors of PKP precursors is indicative of their location and whether they
370 are caused by source or receiver side structure (Haddon & Cleary, 1974). We use PKP data from
371 Thomas et al. (1999) who observe several scatterers beneath Europe and Eastern Asia. Of the
372 data used in Thomas et al. (1999), we focus on a single event occurring on 15 September, 1992
373 which shows clear PKP precursors. We only use data recorded at the Gräfenberg array and not
374 the larger GRSN array to avoid spatial aliasing. In this example, the PKP precursors appear to
375 be coherent from visual inspection of the seismograms (Fig 8. Coherent precursors suggest they
376 probably originate from localised scatterers such as an Ultra Low Velocity Zone (ULVZ) (Ma &
377 Thomas, 2020).

378 Fig 8 shows the traces used for this example and the clusters found by our algorithm. The data
379 have the instrument response removed and are filtered between 0.5 and 2 Hz before the beam-
380 forming process. We used a time window of 10 s before the predicted PKIKP arrival and the same

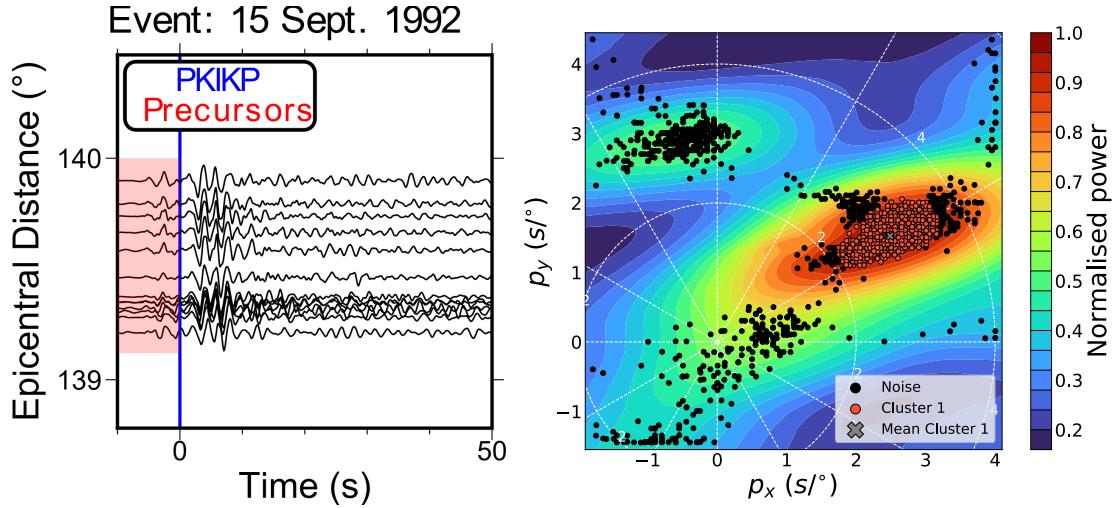


Figure 8. Example application of the method on PKP precursors. This example uses data from the 15 September, 1992 event recorded at the Gräfenberg array in Germany (GR) filtered between 0.5 and 2.0 Hz. The left subfigure shows the traces used in the example which are aligned on the predicted PKIKP arrival time and the time window for the analysis shown in red. On the right, the result of the algorithm with parameters of $\epsilon = 0.2 \text{ s/}^\circ$ and $\text{MinPts} = 0.25$.

381 DBSCAN parameters found from the tuning ($\epsilon = 0.20 \text{ s/}^\circ$ and $\text{MinPts} = 0.25$). The method iden-
 382 tifies a single precursor arriving with a backazimuth of $58.6^\circ \pm 2.3^\circ$ and a horizontal slowness of
 383 $2.93 \text{ s/}^\circ \pm 0.32 \text{ s/}^\circ$. This is similar to the slowness vector properties of the dominant arrival found
 384 by Thomas et al. (1999) arriving 6.5 s before PKIKP with a horizontal slowness of 2.8 s/° and
 385 backazimuth of 53.6° . Unlike Thomas et al. (1999), we only identify one precursor rather than
 386 three. We believe this is because our time window encompasses all precursors meaning if one
 387 precursor has a significantly higher amplitude it may be the only one recovered. Furthermore, vi-
 388 sual inspection of waveforms suggests a single dominant precursor (Fig 8). The range of possible
 389 horizontal slowness of this PKP precursor inferred from the uncertainty of the measurement (2.93
 390 $\text{ s/}^\circ \pm 0.32 \text{ s/}^\circ$) at a distance of approximately 140° means this precursor could originate from
 392 either source side or receiver side structure (Haddon & Cleary, 1974).

393 4.2 Rayleigh wave multipathing

394 The second example shows the identification of multipathed Rayleigh waves. From this obser-
 395 vation, the phase velocities and backazimuths of the multipathed arrivals can be measured and

analysed with uncertainty bounds. Xia et al. (2018) identify multipathing in Rayleigh waves in the western US and suggest this is caused by the transition from continental to coastal to oceanic structure each with unique velocity profiles. We analyse Rayleigh waves from an event on 05 January 2013 recorded at the Southern California Seismic Array (CI) to identify multipathing and hypothesise some potential causes. The instrument response is removed and traces are filtered between 0.04 and 0.06 Hz. The time window used in the relative beamforming is 200 s before and after the predicted arrival time assuming a velocity of 3.5 km/s. In this example, the points in each cluster are distributed over a different slowness-space scale that is an order of magnitude lower than in the body wave examples. The difference is due to the Rayleigh wave velocity and the change in units. p_x/p_y for body waves will vary on the order of 10^0 , whereas for Rayleigh waves p_x/p_y vary on the order of 10^{-1} , an order of magnitude lower. Because of this, the ϵ parameter is also lowered by an order of magnitude from 0.20 s/° found from tuning to 0.02 s/km .

Fig 9 shows the result of the clustering method, which identifies three multipathed arrivals with backazimuths of $319^\circ \pm 0.7^\circ$, $344^\circ \pm 1.3^\circ$ and $299^\circ \pm 1.4^\circ$ and velocities of 3.6 ± 0.025 , 3.5 ± 0.032 and 3.8 ± 0.093 km/s respectively. For each arrival, we mark the path from the mean station location along the mean backazimuth (dashed white line in Fig 9) to determine a possible cause for the multipathing. Also shown are the paths showing the backazimuth uncertainty bounds (solid white lines in Fig 9), which suggest it is reasonable to hypothesise possible causes of the measurements. We investigate dispersion in the wave velocities by repeating the analysis in three frequency bands of $0.035 - 0.045$, $0.045 - 0.055$ and $0.055 - 0.065$ Hz, finding differences in the number of arrivals and their backazimuths, but no absolute slowness variation between frequencies (See Supplementary Figure 1). We argue this is a result of the different scale lengths of the structures which cause the observed multipathing, and not because of a property of the material the wave is traveling through.

The top and middle paths may come from interactions with the boundary between the continental and coastal regions, which agrees with the interpretation of Xia et al. (2018). The direction of the western most arrival suggests it could be caused by interacting with a coastal-ocean velocity transition or possibly due to more localised velocity variations. Further modelling is beyond the

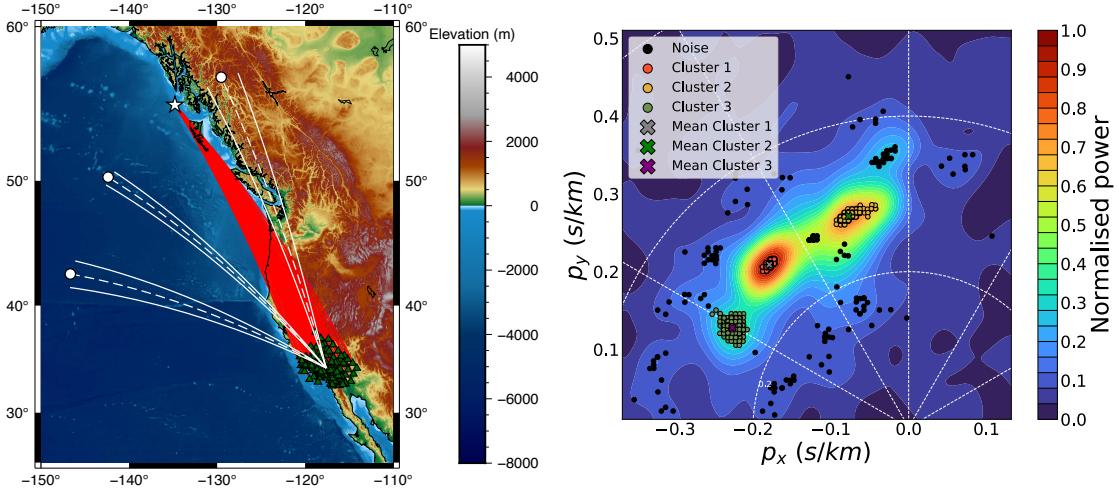


Figure 9. Example application of the method for identifying multipathing in surface waves. The left subfigure shows the raypaths (red lines) from the 05 January, 2013 event (white star) to the Southern California Seismic Array (CI) stations (green triangles). Before the beamforming, the data was filtered between 0.04 and 0.06 Hz. In this example, three arrivals have been identified by the algorithm (right subfigure). For each arrival, a path is marked from the mean station location along the mean backazimuth to a point with the same epicentral distance as the event (dashed white lines and circle). The solid white lines indicate the uncertainty bounds of the backazimuth for the measurement.

424 scope of this work, but our results demonstrate the potential of the method to investigate such
426 phenomena in an efficient way.

427 The phase velocities of the arrivals may be indicative of azimuthal anisotropy beneath the array.
428 The phase velocities of the central and eastern most arrival are the same within the uncertainties
429 (3.6 ± 0.025 km/s and 3.5 ± 0.032 km/s respectively). The western most arrival moves with a
430 significantly higher phase velocity over the array (3.8 ± 0.093 km/s) along a backazimuth of $299^\circ \pm$
431 1.4° . While we do not have enough measurements to fully explore the nature of this azimuthal
432 anisotropy beneath the array, our observation of a faster arrival from 299° is in line with that found
433 by Alvizuri & Tanimoto (2011) who report a fast direction of approximately 290° . Further analysis
434 would be needed to recover the anisotropic properties, but this example shows how our technique
435 can be used to identify statistically significant differences in phase velocity measurements.

436 **5 CODE GUIDELINES**

437 This section outlines some guidance to use this technique in terms of parameter selection and
438 computation time. There are many potential aspects of a study that can influence the method's
439 effectiveness such as frequency bands, array size and configuration or local receiver side structure.
440 The tuning process (Section 3) shows we cover a range of frequency bands (Table 1) and array
441 sizes (10 – 50 stations) and the sub arrays have a wide range of configurations. For applications
442 analysing body waves in similar frequency bands (0.1 – 1 Hz) with a similar array size (10 –
443 50), we recommend the parameters ($\text{MinPts} = 0.25$, $\epsilon = 0.2 \text{ s}^\circ$) used here as a starting point and
444 adjusted if necessary.

445 The number of peaks above the noise threshold should be equal to the maximum number of
446 arrivals of interest or expect to be possible. The noise threshold was determined to be three times
447 the noise estimate through exploratory analysis and found to give satisfactory results, but this can
448 be changed depending on the application. DBSCAN parameters ϵ and MinPts of 0.20 s° and 0.25
449 respectively will work well for identifying single arrivals and is relatively intolerant to noise. If the
450 study is searching for multipathing, changing MinPts to 0.15 and keeping ϵ as 0.20 s° increases
451 the accuracy of the multipathed arrivals from 66 % to 75 % but decreases the accuracy of the noisy
452 arrivals from 80 % to 44 %. These alternative parameters would require visual inspection of those
453 identified as multipathing by the algorithm but would significantly reduce the amount of visual
454 inspection as observations with one arrival need not be visually inspected.

455

456 For surface waves, the algorithm also works well after changing ϵ to 0.02 s/km. For applications with significantly different frequency bands or array size or searching for a very specific phenomenon, the DBSCAN parameters may need to be tuned to optimise performance (Section 3).
457
458 The remaining parameters can be kept the same. Sensible beamforming practice such as avoiding
459 spatial aliasing still applies when using this method.
460

461 The computationally intensive part of the method is the bootstrap sampling and the beam-
462 forming on each sample, which must be performed for each observation; the cluster analysis is
463 comparatively quick. However, the code is trivially parallelisable over observations since each is
464 independent of all the others. The code is written in Python, is easily editable and freely avail-
465 able (https://github.com/eejwa/Array_Seis_Circle). The code has been parallelised so
466 the bootstrap sampling can be spread over several cores and uses Numba (Lam et al., 2015) to
467 compile the functions into machine code before execution. Further improvements in efficiency
468 could be made by rewriting the algorithm in more efficient languages such as Julia, C++ or Fort-
469 ran, and investigating further performance improvements possible with the existing code base.
470 For an example array with 20 stations, a time window of 30 seconds, added[id=JW] sampling rate
471 of 0.05s and searching over a grid of slowness vector properties with 14641 vectors (a grid where
472 each axis covers 6 s/ $^{\circ}$ in increments of 0.05 s/ $^{\circ}$), each bootstrap sample takes approximately 1.6
473 seconds to process. This makes tens of observations viable on a handful of cores such as on a
474 desktop machine. Larger datasets (thousands of observations) can be processed on the order of
475 hours using hundreds of cores.

476 6 CONCLUSIONS

477 Slowness vector measurements have been used to understand a variety of Earth structures and
478 phenomena. They are typically used to identify wavefield perturbations, scattering and event/noise
479 source localisation. While this analysis is a common tool used by seismologists, studies are limited
480 because of the necessary and subjective visual inspection of observations. Interpretation of the
481 measurements is limited by uncertainties such as the contribution of array geometry, noise and

482 local structure. These may result in different slowness vector measurements depending on which
483 stations are used in the analysis.

484 In this study, we described a method to automate slowness vector measurement, estimate the
485 uncertainties and identify the number of possible arrivals. To do this, we bootstrap sample the
486 waveforms and in each sample use a relative beamforming process to measure the coherent power
487 and recover slowness vector properties of potential arrivals. These slowness vector properties are
488 collected and the clustering algorithm DBSCAN is used to identify arrivals. The mean of the clus-
489 ters gives the backazimuth and horizontal slowness and the spread of the cluster gives uncertainty
490 estimates of phenomena which may vary the slowness vector measurement depending on which
491 subset of stations are used. We use a linear beamforming approach but other beamforming methods
492 such as phase weighted stacking (Schimmel & Paulsen, 1997) and cross correlation beamforming
493 (Ruigrok et al., 2017) can be used.

494 We tuned the DBSCAN parameters on a data set with 0, 1 and 2 arrivals and achieved $> 90\%$
495 accuracy in recovering these arrivals. We present examples of analysis of scattered P wave energy
496 and Rayleigh wave multipathing. The advantage this method brings to these applications is the
497 ability to automatically identify the arrivals and measure the slowness vectors with uncertainty
498 estimates. The difference in spatial scale and wavelengths used in these examples shows that our
499 approach is applicable to studying Earth properties at a wide variety of spatial scales. Using this
500 method, it may be possible to analyse slowness vector properties on larger data sets with reduced
501 need for subjective visual inspection. In addition, uncertainties can also be quantified and used
502 alongside the measurements. This technique makes 1000s of observations feasible in a matter of
503 hours and allows for global-scale slowness vector observations to be made.

504 **ACKNOWLEDGMENTS**

505 We thank reviewers Steve Gibbons and Yu Gu for their constructive reviews which has improved
506 this manuscript. ObspyDMT (Hosseini & Sigloch, 2017) was used to download data. GMT (Wes-
507 sel et al., 2013) was used to make some of the figures. Predictions from 1D velocity models were
508 made using the TauP toolkit Crotwell et al. (1999). We thank the Department of Geology and Geo-

Automatic Slowness Vector Measurements of Seismic Arrivals with Uncertainty Estimates using Unsupervised Learning

physics, University of Utah for hosting a collaborative visit and the Centre for High Performance Computing (CHPC) for computer resources and support. JW is supported by the NERC DTP Spheres grant NE/L002574/1, AN was funded by NERC standard grant NE/R001154/1 (REMIS: Reliable Earthquake Magnitudes for Induced Seismicity) and MT was partially supported by NSF grant EAR-1723081.

DATA AVAILABILITY

Code to perform the analysis is available at: https://github.com/eejwa/Array_Seis_Circle. Data used for tuning and the examples is available to download from: <https://figshare.com/s/fbcb167ad15d581cf4e>. Seismic arrays used were the Kaapvaal array (James et al., 2001), the Gräfenberg array (Federal Institute For Geosciences And Natural Resources (BGR), 1976) [<https://doi.org/10.25928/mbx6-hr74>] and the Southern California Seismic Network [<https://doi.org/10.7914/SN/CI>] (California Institute of Technology and United States Geological Survey Pasadena, 1926).

References

- Alvizuri, C. & Tanimoto, T., 2011. Azimuthal anisotropy from array analysis of Rayleigh waves in Southern California, *Geophysical Journal International*, **186**(3), 1135–1151.
- Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J., 1999. OPTICS: ordering points to identify the clustering structure, *ACM Sigmod record*, **28**(2), 49–60.
- Bear, L. K. & Pavlis, G. L., 1997. Estimation of slowness vectors and their uncertainties using multi-wavelet seismic array processing, *Bulletin of the Seismological Society of America*, **87**(3), 755–769.
- Behr, Y., Townend, J., Bowen, M., Carter, L., Gorman, R., Brooks, L., & Bannister, S., 2013. Source directionality of ambient seismic noise inferred from three-component beamforming, *Journal of Geophysical Research: Solid Earth*, **118**(1), 240–248.

- 534 Bentham, H. & Rost, S., 2014. Scattering beneath Western Pacific subduction zones: evidence
535 for oceanic crust in the mid-mantle, *Geophysical Journal International*, **197**(3), 1627–1641.
- 536 Berteussen, K., 1976. The origin of slowness and azimuth anomalies at large arrays, *Bulletin of
537 the Seismological Society of America*, **66**(3), 719–741.
- 538 Bondár, I., North, R. G., & Beall, G., 1999. Teleseismic slowness-azimuth station corrections for
539 the International Monitoring System seismic network, *Bulletin of the Seismological Society of
540 America*, **89**(4), 989–1003.
- 541 Bower, D. J., Gurnis, M., & Seton, M., 2013. Lower mantle structure from paleogeographically
542 constrained dynamic Earth models, *Geochemistry, Geophysics, Geosystems*, **14**(1), 44–63.
- 543 Bowers, D. & Selby, N. D., 2009. Forensic seismology and the Comprehensive Nuclear-Test-Ban
544 Treaty, *Annual Review of Earth and Planetary Sciences*, **37**, 209–236.
- 545 California Institute of Technology and United States Geological Survey Pasadena, 1926. South-
546 ern California seismic network, *International Federation of Digital Seismograph Networks,
547 Dataset/Seismic Network*.
- 548 Campello, R. J., Moulavi, D., & Sander, J., 2013. Density-based clustering based on hierarchical
549 density estimates, in *Pacific-Asia conference on knowledge discovery and data mining*, pp. 160–
550 172, Springer.
- 551 Campello, R. J., Moulavi, D., Zimek, A., & Sander, J., 2015. Hierarchical density estimates for
552 data clustering, visualization, and outlier detection, *ACM Transactions on Knowledge Discovery
553 from Data (TKDD)*, **10**(1), 1–51.
- 554 Cao, A. & Romanowicz, B., 2007. Locating scatterers in the mantle using array analysis of PKP
555 precursors from an earthquake doublet, *Earth and Planetary Science Letters*, **255**(1-2), 22–31.
- 556 Celebi, M. E., Kingravi, H. A., & Vela, P. A., 2013. A comparative study of efficient initialization
557 methods for the k-means clustering algorithm, *Expert systems with applications*, **40**(1), 200–
558 210.
- 559 Chevrot, S., Sylvander, M., Benahmed, S., Ponsolles, C., Lefevre, J., & Paradis, D., 2007. Source
560 locations of secondary microseisms in western Europe: Evidence for both coastal and pelagic
561 sources, *Journal of Geophysical Research: Solid Earth*, **112**(B11).

- Automatic Slowness Vector Measurements of Seismic Arrivals with Uncertainty Estimates using Unsupervised Learning*
- 562 Cottaar, S. & Romanowicz, B., 2012. An unusually large ULVZ at the base of the mantle near
Hawaii, *Earth and Planetary Science Letters*, **355**, 213–222.
- 563 Crotwell, H. P., Owens, T. J., & Ritsema, J., 1999. The TauP Toolkit: Flexible seismic travel-time
and ray-path utilities, *Seismological Research Letters*, **70**(2), 154–160.
- 564 Efron, B., 1992. Bootstrap methods: another look at the jackknife, in *Breakthroughs in statistics*,
pp. 569–593, Springer.
- 565 Ertöz, L., Steinbach, M., & Kumar, V., 2003. Finding clusters of different sizes, shapes, and den-
sities in noisy, high dimensional data, in *Proceedings of the 2003 SIAM international conference*
on data mining, pp. 47–58, SIAM.
- 566 Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discover-
ing clusters in large spatial databases with noise., in *Kdd*, vol. 96, pp. 226–231.
- 567 Federal Institute For Geosciences And Natural Resources (BGR), 1976. German Regional Seis-
mic Network (GRSN).
- 568 Frost, D. A., Rost, S., Selby, N. D., & Stuart, G. W., 2013. Detection of a tall ridge at the
core–mantle boundary from scattered PKP energy, *Geophysical Journal International*, **195**(1),
558–574.
- 569 Gal, M., Reading, A., Ellingsen, S., Koper, K., Gibbons, S., & Näsholm, S., 2014. Improved
implementation of the fk and Capon methods for array analysis of seismic noise, *Geophysical*
Journal International, **198**(2), 1045–1054.
- 570 Gal, M., Reading, A., Ellingsen, S., Koper, K., Burlacu, R., & Gibbons, S., 2016. Deconvolution
enhanced direction of arrival estimation using one-and three-component seismic arrays applied
to ocean induced microseisms, *Geophysical Journal International*, **206**(1), 345–359.
- 571 Gerstoft, P., Fehler, M. C., & Sabra, K. G., 2006. When katrina hit california, *Geophysical*
Research Letters, **33**(17).
- 572 Gerstoft, P., Shearer, P. M., Harmon, N., & Zhang, J., 2008. Global P, PP, and PKP wave micro-
seisms observed from distant storms, *Geophysical Research Letters*, **35**(23).
- 573 Gibbons, S. J. & Ringdal, F., 2011. Seismic monitoring of the North Korea nuclear test site
using a multichannel correlation detector, *IEEE transactions on geoscience and remote sensing*,

- 575 **50**(5), 1897–1909.
- Gibbons, S. J., Ringdal, F., & Kværna, T., 2008. Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays, *Geophysical Journal International*, **172**(1), 405–421.
- 576 Gibbons, S. J., Schweitzer, J., Ringdal, F., Kværna, T., Mykkeltveit, S., & Paulsen, B., 2011. Improvements to seismic monitoring of the European Arctic using three-component array processing at SPITS, *Bulletin of the Seismological Society of America*, **101**(6), 2737–2754.
- 577 Gibbons, S. J., Näsholm, S., Ruigrok, E., & Kværna, T., 2018. Improving slowness estimate stability and visualization using limited sensor pair correlation on seismic arrays, *Geophysical Journal International*, **213**(1), 447–460.
- 578 Haddon, R. & Cleary, J., 1974. Evidence for scattering of seismic PKP waves near the mantle-core boundary, *Physics of the Earth and Planetary Interiors*, **8**(3), 211–234.
- 579 Hosseini, K. & Sigloch, K., 2017. obspyDMT: a Python toolbox for retrieving and processing of large seismological datasets, *Solid Earth*, **8**.
- 580 James, D., Fouch, M., VanDecar, J., Van Der Lee, S., & Group, K. S., 2001. Tectospheric structure beneath southern Africa, *Geophysical research letters*, **28**(13), 2485–2488.
- 581 Ji, C., Tsuboi, S., Komatitsch, D., & Tromp, J., 2005. Rayleigh-wave multipathing along the west coast of North America, *Bulletin of the Seismological Society of America*, **95**(6), 2115–2124.
- 582 Kim, D., Lekić, V., Ménard, B., Baron, D., & Taghizadeh-Popp, M., 2020. Sequencing seismograms: A panoptic view of scattering in the core-mantle boundary region, *Science*, **368**(6496), 1223–1228.
- 583 Koch, K. & Kradolfer, U., 1999. Determination of mislocation vectors to evaluate bias at GSETT-3 primary stations, *Journal of seismology*, **3**(2), 139–151.
- 584 Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P., 2019. Machine learning in seismology: Turning data into insights, *Seismological Research Letters*, **90**(1), 3–14.
- Korenaga, J., 2013. Stacking with dual bootstrap resampling, *Geophysical Journal International*, **195**(3), 2023–2036.
- 586 Lam, S. K., Pitrou, A., & Seibert, S., 2015. Numba: A llvm-based python jit compiler, in *Pro-*

- Automatic Slowness Vector Measurements of Seismic Arrivals with Uncertainty Estimates using Unsupervised Learning
587 Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, pp. 1–6.
- Landès, M., Hubans, F., Shapiro, N. M., Paul, A., & Campillo, M., 2010. Origin of deep ocean microseisms by using teleseismic body waves, *Journal of Geophysical Research: Solid Earth*, **115**(B5).
- Lin, C.-H. & Roecker, S., 1996. P-wave backazimuth anomalies observed by a small-aperture seismic array at Pinyon Flat, southern California: Implications for structure and source location, *Bulletin of the Seismological Society of America*, **86**(2), 470–476.
- Liu, Q., Koper, K. D., Burlacu, R., Ni, S., Wang, F., Zou, C., Wei, Y., Gal, M., & Reading, A. M., 2016. Source locations of teleseismic P, SV, and SH waves observed in microseisms recorded by a large aperture seismic array in China, *Earth and Planetary Science Letters*, **449**, 39–47.
- Ma, X. & Thomas, C., 2020. Small-scale scattering heterogeneities in the lowermost mantle from a global analysis of PKP precursors, *Journal of Geophysical Research: Solid Earth*, **125**(3), e2019JB018736.
- MacQueen, J. et al., 1967. Some methods for classification and analysis of multivariate observations, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA.
- Maupin, V., 2011. Upper-mantle structure in southern Norway from beamforming of Rayleigh wave data presenting multipathing, *Geophysical Journal International*, **185**(2), 985–1002.
- McInnes, L. & Healy, J., 2017. Accelerated hierarchical density based clustering, in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 33–42, IEEE.
- Niu, F. & Kawakatsu, H., 1997. Depth variation of the mid-mantle seismic discontinuity, *Geophysical Research Letters*, **24**(4), 429–432.
- Ogden, C., Bastow, I., Gilligan, A., & Rondenay, S., 2019. A reappraisal of the H- κ stacking technique: implications for global crustal structure, *Geophysical Journal International*, **219**(3), 1491–1513.
- Ritsema, J., Kaneshima, S., & Haugland, S. M., 2020. The dimensions of scatterers in the lower mantle using USArray recordings of S-wave to P-wave conversions, *Physics of the Earth and Planetary Interiors*, **306**, 106541.

- Rost, S. & Thomas, C., 2002. Array seismology: Methods and applications, *Reviews of geophysics*, **40**(3).
- Rost, S. & Thomas, C., 2009. Improving seismic resolution through array processing techniques, *Surveys in Geophysics*, **30**(4-5), 271–299.
- Roux, P. & Ben-Zion, Y., 2017. Rayleigh phase velocities in Southern California from beam-forming short-duration ambient noise, *Geophysical Journal International*, **211**(1), 450–454.
- Ruigrok, E., Gibbons, S., & Wapenaar, K., 2017. Cross-correlation beamforming, *Journal of Seismology*, **21**(3), 495–508.
- Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophysical Journal International*, **130**(2), 497–505.
- Schmidt, R., 1986. Multiple emitter location and signal parameter estimation, *IEEE transactions on antennas and propagation*, **34**(3), 276–280.
- Schumacher, L. & Thomas, C., 2016. Detecting lower-mantle slabs beneath Asia and the Aleutians, *Geophysical Journal International*, **205**(3), 1512–1524.
- Schweitzer, J., 2001. Slowness corrections—One way to improve IDC products, *pure and applied geophysics*, **158**(1), 375–396.
- Selby, N. D., 2011. Improved teleseismic signal detection at small-aperture arrays, *Bulletin of the Seismological Society of America*, **101**(4), 1563–1575.
- Stockmann, F., Cobden, L., Deschamps, F., Fichtner, A., & Thomas, C., 2019. Investigating the seismic structure and visibility of dynamic plume models with seismic array methods, *Geophysical Journal International*, **219**(Supplement_1), S167–S194.
- Teanby, N., Kendall, J.-M., & Van der Baan, M., 2004. Automation of shear-wave splitting measurements using cluster analysis, *Bulletin of the Seismological Society of America*, **94**(2), 453–463.
- Thomas, C., Weber, M., Wicks, C., & Scherbaum, F., 1999. Small scatterers in the lower mantle observed at German broadband arrays, *Journal of Geophysical Research: Solid Earth*, **104**(B7), 15073–15088.

Automatic Slowness Vector Measurements of Seismic Arrivals with Uncertainty Estimates using Unsupervised Learning

- Thomas, C., Kendall, J.-M., & Weber, M., 2002. The lowermost mantle beneath northern Asia—I. Multi-azimuth studies of a D heterogeneity, *Geophysical Journal International*, **151**(1), 279–295.
- Thorne, M. S., Pachhai, S., Leng, K., Wicks, J. K., & Nissen-Meyer, T., 2020. New Candidate Ultralow-Velocity Zone Locations from Highly Anomalous SPdKS Waveforms, *Minerals*, **10**(3), 211.
- Valentine, A. P. & Trampert, J., 2012. Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data, *Geophysical Journal International*, **189**(2), 1183–1202.
- Valentine, A. P. & Woodhouse, J. H., 2010. Approaches to automated data selection for global seismic tomography, *Geophysical Journal International*, **182**(2), 1001–1012.
- Ward, J., Nowacki, A., & Rost, S., 2020. Lateral Velocity Gradients in the African Lower Mantle Inferred From Slowness Space Observations of Multipathing, *Geochemistry, Geophysics, Geosystems*, **21**(8), e2020GC009025.
- Wessel, P., Smith, W. H., Scharroo, R., Luis, J., & Wobbe, F., 2013. Generic mapping tools: improved version released, *Eos, Transactions American Geophysical Union*, **94**(45), 409–410.
- Xia, Y., Ni, S., & Tape, C., 2018. Multipathing Rayleigh Waves From Long-Distance Noise Cross Correlation Along an Ocean-Continent Boundary (Alaska to California), *Geophysical Research Letters*, **45**(12), 6051–6060.
- Yang, Z. & He, X., 2015. Oceanic crust in the mid-mantle beneath west-central Pacific subduction zones: evidence from S to P converted waveforms, *Geophysical Journal International*, **203**(1), 541–547.