Geophysical Journal International

Geophys. J. Int. (2022) **231,** 2156–2172
Advance Access publication 2022 August 23
GJI Seismology

https://doi.org/10.1093/gji/ggac322

The use of weighted self-organizing maps to interrogate large seismic data sets

S. G. Meyer [1,2] A. M. Reading [2] and A. P. Bassom [2]

[1] *Institute of Mine Seismology,* 19 *Jeanine Street, Sudbury, ON P3B 0E5, Canada. E-mail: stephen.meyer@imseismology.org*
[2] *School of Natural Sciences, University of Tasmania, Private Bag* 37*, Hobart, Tasmania* 7001*, Australia*

Accepted 2022 August 17. Received 2022 August 2; in original form 2022 January 5

SUMMARY

Modern microseismic monitoring systems can generate extremely large data sets with signals originating from a variety of natural and anthropogenic sources. These data sets may contain multiple signal types that require classification, analysis and interpretation: a considerable task if done manually. Machine learning techniques may be applied to these data sets to expedite and improve such analysis. In this study, we apply an unsupervised technique, the Self-Organizing Map (SOM), to high-volume data recorded by an in-mine microseismic network. This represents a good example of a large seismic data set that contains a wide range of signals, owing to the diversity of source processes occurring within the mine. The signals are quantified by extracting a number of features (temporal and spectral) from the waveforms which are provided as input data for the SOM. We develop and implement a weighted variant of the SOM in which the contributions of various different features to the training of the map are allowed to evolve. The standard and weighted SOMs are applied to the data, and the output maps compared. Both variants are able to separate source types based on the waveform characteristics, allowing for rapid, automatic classification of signals and the ability to find sources with similar waveforms. Fast classification of such signals provides practical benefit by automatically discarding waveforms associated with anthropogenic sources within the mine while seismic signals originating from genuine microseismic events, which constitute a small fraction of all signals, can be prioritized for subsequent processing and analysis. The weighted variant provides an exploratory tool through quantification of the contribution of different features to the clustering process. This helps to optimize the performance of the SOM through the identification of redundant features. Furthermore, those features that are assigned large weights are considered to be more representative of the source generation processes as they contribute more to the cluster separation process. We apply weighted SOMs to data from a mine recorded during two different time periods, corresponding to different stages of the mine development. Changes in feature importance and in the observed distribution of feature values indicate evolving source generation processes and may be used to support investigatory analysis. The weighted SOM therefore represents an effective tool to help manage and investigate large seismic data sets, providing both practical benefit and insight into underlying event mechanisms.

Key words: Earthquake source observations; Induced seismicity; Seismic noise; Machine learning.

1 INTRODUCTION

Seismology, and its practical application, starts with the recording of ground motions by sensors as the seismic waves propagate through the sensor locations. These ground motions and their characteristics are dependent on the source processes that generate them as well as the propagation path taken from source to receiver. Studying ground motion waveform records, and extracting information from them, is a core principle of observational seismology and either focuses on the source or the path (Aki & Richards 2002). The volume of waveform catalogues has increased due to improvements in sensor technology, and hence the ease of deploying dense

seismic networks. In parallel, the limited capacity for manual analysis by human experts demands automated tools to manage these data sets and extract the desired information. The objective of such automation may be highly specific, such as the identification of seismic waves originating from large, hazardous seismic events, so that these events may be located and their magnitude and moment tensor determined (Ekström *et al*. 2012). However, tools that assist with more investigatory analyses are also required.

Seismic sources can be considered as forces, acting within the rock or substrate, that generate seismic waves. The nature of these radiated waves can be considered representative of those sources. Waveform characteristics are dependent on many factors, such as deformation type (e.g. shear, volume change or impact), deformation rate and amplitude, and the confining stresses at source. In cases where the propagation path may be assumed to be constant (or effectively so), any variations in the observed waveforms must be due to changes in the source processes generating the waves. To utilize these waveforms and changes in their characteristics, the raw ground motion records need to be converted into quantifiable values that describe the waveforms, often based on aspects related to the amplitude, frequency content and other statistical parameters. This may be achieved through feature (or attribute) extraction or calculation (Provost *et al*. 2017).

Representing a waveform data set as a set of representative features opens up a rich set of options in machine learning or pattern recognition techniques. Large data sets with diverse feature content are, in general, well suited to machine learning techniques, which can broadly be split into two categories, supervised and unsupervised (Marsland 2015); the defining difference is the presence, or lack, of data with labels. A significant body of ongoing research is focused on using supervised machine learning in the classification of seismic data (e.g. Langer *et al*. 2006; Kortström *et al*. 2016; Provost *et al*. 2017; Reynen & Audet 2017; Tibi *et al*. 2019) and the identification of phase arrivals and locating events (e.g. Zhu & Beroza 2019; Zhu *et al*. 2019; Kriegerowski *et al*. 2019). The training sets used by the given studies make use of labels that have been previously assigned by human experts. Supervised learning techniques are therefore used to carry out a repetitive task in the manner of a human analyst or make some testable predictions using aspects of the data that may not be initially apparent (e.g. Rouet-Leduc *et al*. 2017).

While supervised learning techniques extend the work that can be performed by a human, unsupervised techniques provide tools to help domain experts interrogate unwieldy data sets and hence apply their knowledge in a more effective manner. This can often take the form of an algorithm being applied to suggest patterns or similarities in data that are not readily apparent, hence illuminating this information. Unsupervised techniques provide a method of clustering unlabelled data so that inspection enables the subsequent classification, or application of labels, to the data. For example, Köhler *et al*. (2010) applied a Self-Organizing Map (SOM) to continuous recordings from an active volcano, identifying and labelling signals associated with rockfall and volcano-tectonic sources. In this case, the SOM was able to find patterns in the waveform records and subsequently group similar signals together after the map was trained. The concept of 'training' is often used in machine learning. In the case of supervised techniques, where the training data are labelled, it refers to the method learning how to apply these labels to new data. For unsupervised techniques, where the data are unlabelled, training refers to the identification of similar groups or clusters in a representative data set by the algorithm, with these clusters (e.g. a trained map) being used for the subsequent classification of new data.

Since the original implementation of the SOM by Kohonen (1990), the method has been used widely, in a variety of fields, including seismology (e.g. Köhler & Ohrnberger 2008; Esposito *et al*. 2008; Köhler *et al*. 2010; Ida & Ishida 2022). There have also been a number of developments, improvements and modifications to the original SOM framework. Many of these adaptations are discussed within a theoretical framework and demonstrated on small or synthetic data sets, without thorough analysis of the results. Work by de Bodt & Cottrell (2000) proposed bootstrapping a SOM in order to test the true statistical significance of SOM neighbours while Guérif *et al*. (2005) introduced the μ-SOM, a weighted, iterative method in which the data dimensions have weights applied to them. The evaluation of the weights is computed using a second SOM, trained on features extracted from the component planes of the first SOM.

One of the aims of a SOM is to find informative associations between features that may not be apparent to the human eye. Therefore, making use of all available features would seem to be the ideal, but dealing with data containing an unnecessarily high number of dimensions poses some challenges. First, and most obviously, higher dimensional data sets will require additional computational time to process with likely minimal added benefit in terms of the resulting map quality or insight. Further, strong correlations between certain features will be highlighted, potentially obscuring relations between other feature pairs or groups that are less clear but perhaps more meaningful. As the dimension increases, the 'curse of dimensionality' (Marsland 2015) means that the Euclidean distance becomes a less appropriate measure of the similarity of two samples. Therefore, it is desirable to reduce the number of features used while minimizing any negative impact of this reduction on the clustering results.

In this study, we demonstrate how a SOM can be used to manage and interrogate large seismic data sets containing signals originating from diverse sources, and hence facilitate the investigation of the forces behind the generating sources. A variant of the SOM known as the weighted SOM (ω-SOM) is described in which the different features are assigned weights based on their relative contribution to the clustering procedure. The performance of the ω-SOM relative to the standard SOM is assessed using synthetic data and two real, large-volume data sets, separated in time. The results of the unsupervised learning are appraised for their likely usefulness in a practical setting. This enables us to assess how such techniques might provide benefit in terms of time and cost savings. Finally, the use of the ω-SOM enables greater insight into the features that are characteristic in describing a data set and provides a means of identification of changes to the generative source processes. This is particularly useful when one considers that it is an unsupervised technique, allowing it to be applied easily on new, unlabelled data sets. The method is therefore more powerful than the standard SOM and easier and faster to apply than more complex, supervised methods.

2 DATA

We analyse seismic data recorded by a microseismic monitoring network installed in an underground hard-rock mine in Australia. The in-mine instrumentation comprises an extensive array of three-component geophones with natural frequencies of 4.5 or 14 Hz. Event detection is based on an STA/LTA triggering and association

scheme (Mendecki 1997). For the time period used in this study, the number of operational sensors in the mine varied from 60 to 72, with waveforms being sampled at 6000 Hz.

The monitoring network recorded 280 000 event entries in the database over a 12-month period. Each entry represents a set of triggered and associated seismograms, many of which are genuine microseismic events associated with the brittle failure of the rockmass due to mining induced stress changes. A much larger proportion of events, however, relate to anthropogenic sources, such as machinery associated with the crushing and transport of ore, and blasting activities. This has been confirmed through conventional mine seismology studies: analyses that examine the source location of these signals, source type (i.e. they are better described by single-forces rather than moment tensors), and by the mine providing time and location data on the various activities occurring in the mine (e.g. rockbreakers, crushers and hydraulic hammering). Anthropogenic sources are generally not of interest and their presence in the seismic catalogue can impede seismic hazard assessments, which poses a real and constant challenge for seismically active mines. A sudden flurry of these noise events can also trigger evacuation and exclusion procedures at the mine if the sources are not correctly classified.

3 METHODS

3.1 Feature extraction

The seismic waveform event database (as above) is first processed to extract or calculate waveform features. Some features (Table 1) are based on straightforward, time-domain waveform characteristics (e.g. maximum amplitude, ratio of mean value to maximum amplitude), while others are based on aspects of the waveform that require calculation, such as spectral content (e.g. peak frequency of the FFT spectrum, frequency band amplitude ratios), and statistical measures (e.g. variance of the raw-time domain amplitude) following the first part of the study by Provost *et al*. (2017). Each feature is calculated based on a single waveform record with multiple waveform records being associated with each event. The average of each feature across a number of seismograms (in this study, 6) with the highest recorded maximum amplitude is taken and used to represent the event in question. The features are normalized to have values between 0 and 1 prior to training the SOM.

Fig. 1 shows examples of 4 different event types in the database together with the feature value for that event type superposed on the histograms of distributions of each feature (Table 1) for the entire data set. The first waveform (Event 1, blue) has high frequency, impulsive arrivals with medium amplitude and is associated with a $m_W = -1.5$ fault slip microseismic event (where m_W is the moment magnitude). The second waveform (Event 2, red) is associated with a typical blasting sequence for extending underground tunnels. The third waveform (Event 3, green) is from a low stress, $m_W = 0.0$ event associated with the expansion of the block cave in the mine and has fairly low amplitude and low frequency waveforms. The final waveform (Event 4, orange) has very low amplitude and frequency waveforms and is associated with machinery in the mine. While seismic waves associated with events types similar to Event 1 (microseismic event) and Event 3 (block cave expansion) are of the most interest, they often constitute a very small portion of the seismic waves recorded.

3.2 A standard SOM

Originally proposed by Kohonen (1990), the SOM has become a popular method of unsupervised classification to visualize and interpret large, high-dimensional data sets. It functions by mapping a large set of high dimensional data to a lower dimensional set of vectors (the map, usually 2-D). An attractive aspect of the SOM is the preservation of the internal topology of the data structure.

If $\{\mathbf{x} : \mathbf{x_i} \in \mathbb{R}^N, i = 1, 2, ..., K\}$ represents the input, N-dimensional data set, we first generate prototype vectors $\{\mathbf{m} : \mathbf{m}_j \in \mathbb{R}^N, j = 1, 2, ..., M\}$, where M is determined from the size of the map. These prototype vectors are often referred to as weight vectors in other SOM studies but here we shall adopt the term 'prototype' to avoid confusion, as we reserve the word 'weight' in the context of feature weights. We initialize the prototype vectors randomly, although selecting an appropriate initial state may lead to faster convergence (Kohonen 2013).

The input data vectors remain fixed while the prototype vectors evolve during training so that $\mathbf{m} = \mathbf{m}(t)$, where t is the training iteration. At each training step, an input data vector $\mathbf{x}_i$ is randomly selected. The prototype vector, $\mathbf{m}^b$, that is nearest to x_i in N-dimensional space so that

$$||\mathbf{x}_i - \mathbf{m}^b|| = \min_j\{||\mathbf{x}_i - \mathbf{m}_j||\}. \tag{1}$$

is selected as the corresponding best matching unit (BMU). The prototype vectors are updated according to the applied neighbourhood function:

$$\mathbf{m_j}(t+1) = \mathbf{m_j}(t) + \alpha(t)h_{bj}(t)[\mathbf{x_i} - \mathbf{m_j}(t)], \tag{2}$$

where t is the iteration, $\alpha(t)$ is the learning rate and$h_{bj}(t)$ is the neighbourhood function centred on the BMU:

$$h_{bj}(t) = \exp\left(-\frac{||\mathbf{r}_b - \mathbf{r}_j||}{2\sigma^2(t)}\right). \tag{3}$$

In this expression, $\mathbf{r}_b$ and $\mathbf{r}_j$ are the positions on the SOM grid of neurons b and j, and σ is the size of the neighbourhood search radius, which is represented by a Gaussian kernel.

The learning rate, $\alpha(t)$, and neighbourhood function radius, $\sigma(t)$ do not need to follow a strict formulae. It is, however, critical that they are monotonically decreasing functions that begin with a large value and become much smaller (Kohonen 2013). Further, these functions should initially change very slowly while the algorithm carries out 'large scale' or 'rough' sorting. A number of formulations that describe how these parameters evolve have been used. We adopt the formulation used by Chaudhary *et al*. (2014) in which

$$\alpha(t) = \alpha(0)\left(\frac{\alpha(T)}{\alpha(0)}\right)^{t/T} \tag{4}$$

while the search radius $\sigma(t)$ is given by

$$\sigma(t) = \sigma(0)\left(\frac{\sigma(T)}{\sigma(0)}\right)^{t/T}, \tag{5}$$

where T is the total number of training iterations. There is nothing particularly noteworthy of this form except that it meets the fundamental criteria of being monotonically decreasing, and has been shown to be successful in training of SOMs. One attraction is that it is simple to implement, requiring only the initial and final values for α and σ.

The size of the map and the number of training iterations required can vary depending on the complexity and volume of data. While

Table 1. Description of features (attributes) used. Values are calculated for each seismogram of interest, and the six seismograms with the highest recorded maximum amplitude are averaged to represent the event.

#	Name	Description
1	Lg10MaxAmp	Logarithm of the maximum amplitude of the seismogram
2	Lg10MeanR	Logarithm of the ratio of the maximum amplitude to the mean value of the seismogram
3	Lg10MedianR	Logarithm of the ratio of the maximum amplitude to the median value of the seismogram
4	Lg10Var	Logarithm of the variance of the amplitude of the seismogram
5	Lg10Kurtosis	Kurtosis of the amplitude of the seismogram
6	Lg10AutoCor10	Logarithm of the number of peaks in the autocorrelation function using a width of 10 samples
7	Lg10AutoCor100	Number of peaks in the autocorrelation function using a width of 100 samples
8	Lg10N75Max	Logarithm of the portion of the seismogram above 75 per cent of the maximum amplitude
9	Lg10N50Max	Logarithm of the portion of the seismogram above 50 per cent of the maximum amplitude
10	Lg10N25Max	Logarithm of the portion of the seismogram above 25 per cent of the maximum amplitude
11	Lg10ForwardRec	Logarithm of the number of times a new maximum in the absolute value of the amplitude is counted moving forward in time from the first sample in the seismogram
12	Lg10BackwardRec	Logarithm of the number of times a new maximum in the absolute value of the amplitude is counted moving backward in time from the last sample in the seismogram
13	Lg10FFTPeak	Logarithm of the peak frequency of the FFT of the seismogram
14	Lg10FFTVar	Logarithm of the variance of the FFT spectrum
15	Log10FFTKurtosis	Logarithm of the kurtosis of the FFT spectrum
16	FFTR_10-30_300-500	Ratio of FFT amplitude between 10–30 Hz and 300–500 Hz
17	FFTR_40-60_600-800	Ratio of FFT amplitude between 40–60 Hz and 600–800 Hz
18	Lg10CAD	Logarithm of the cumulative absolute displacement (Mendecki 2019) of the seismogram
19	T90	Duration of seismogram accounting for 90 per cent of total energy

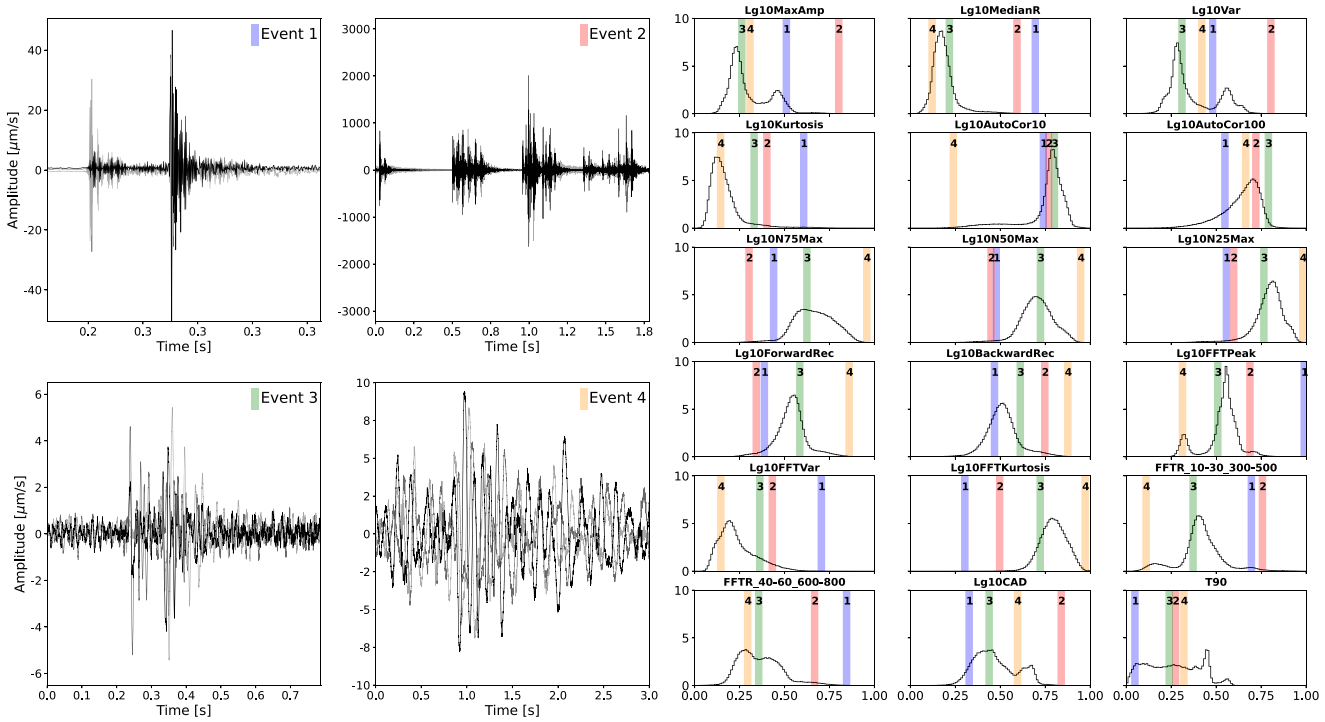

Figure 1. Examples of waveforms and waveform features of the input data set for 4 event types, numbered and coloured blue, red, green and orange for identification in the feature distributions to the right. Seismic records for event types 1–4, grey scale waveforms (black, dark grey, light grey, arbitrarily oriented *X*, *Y* and *Z*). Smaller plots: feature distributions, with colour bars and numbers representing the normalized value of the different features for each of the four event types relative to the full distribution histogram (stepped black line, *X*-axis bin sizes are 0.02, *Y*-axis represents the relative frequency). Features match those in Table 1, except that Feature 2 is omitted for convenience of the illustration.

there are no strict criteria that need to be met, some general guidelines have been established. Each cell of the map should have an average of 50 data entries associated with it after training (Kohonen 2013), and setting the number of cells equal to $5\sqrt{K}$ is reasonable (where K is the total number of input data points, Kohonen 2001). The number of iterations should be equal to roughly 500 times the number of cells (Kohonen 1990). These guidelines enable users to choose sensible initial parameters for SOMs but these can be modified with care.

Once these initialization parameters are set, the training of the SOM is accomplished via the following steps:

Step 1: Select a random data vector and find the corresponding Best Matching Unit (BMU).

Step 2: Update the prototype vectors accordingly, using the BMU and neighbourhood function (eq. 2).

Training of SOMs can be achieved in either batch or step mode (Kohonen 2013), with the latter being used in the original formulation and that presented above. While both approaches can be successful, batch mode has the advantage of being faster, insensitive to training parameters such as the learning rate, α, and deterministic in that training on the same data set with the same parameters will yield a consistent map (Fort *et al*. 2002). On the other hand, while the batch mode can be very sensitive to the initialization of the map, typically the step mode is not. We therefore choose to use the step mode as avoiding complications associated with initialization is more important for our study and outweighs the advantages provided by the batch mode.

The reliance of step mode on the learning rate and its evolution during training is also considered to be manageable as long as this parameter decreases monotonically at a reasonable rate, i.e. the map will be able to be trained (Kohonen 1990). While a precisely repeatable process in which the exact same outcome is reached (as in batch mode) may be desirable in some situations, it is not a vital requirement for this study. In this study, the trained maps are saved and used for all subsequent analysis, so the ability to recreate an identical outcome does not deliver any benefit. That is a repeat map with exactly the same configuration is not required as the maps will not be retrained (at least not with exactly the same input data and parameters).

3.3 A weighted SOM

One of the drawbacks of the SOM is the inability to identify automatically, and hence remove, irrelevant features (Ahmad & Starkey 2018). This is of most concern for data sets that are extremely large or have a very high number of features (dimensions), when efficient simplification of the data through removal of redundant features is strongly desired. Guérif & Bennani (2007) proposed a solution to this through a modification of the SOM, known as the weighted self-organizing map, or ω-SOM. In this approach, each feature is also assigned a certain weight, which represents the respective contribution of that feature to the overall SOM cost function. Their implementation of the weighted SOM is part of a feature selection procedure and was demonstrated from a theoretical optimisation perspective on a couple of synthetic data sets. However the potential insight gained from feature weights applied to some real data was not explored. Within the weighted SOM, features with a lower weight can be considered as contributing less to the training of the SOM while features with larger weights are more useful. An implementation of a weighting procedure is described in the context of a weighted k-means algorithm by Huang *et al*. (2005). Applying this to the SOM framework is similar in concept to the k-means algorithm with the addition of the defining neighbourhood function (eq. 3).

We now introduce the SOM weights, ω_n with $\sum_n \omega_n = 1$ and $\omega_n \geq 0$ for all n, where $n = 1, 2, ..., N$. The standard SOM implementation used Euclidean distance so that the distance d between vectors $\mathbf{x}$ and $\mathbf{m}$ is defined to be

$$d^2(\mathbf{x}, \mathbf{m}) = \sum_n (x_n - m_n)^2, \tag{6}$$

with weighting on a per-dimension basis giving:

$$d^2_\omega(\mathbf{x}, \mathbf{m}) = \sum_n \omega_n^\beta (x_n - m_n)^2, \tag{7}$$

where β is a prescribed constant parameter, discussed in more detail below.

At the start of training, the weights are either randomized or set to be equal; we adopt the latter approach. As with the standard SOM, the first two steps are to: (1) select a data vector at random and identify the best matching unit (BMU) with the distance metric being found using the weighted variant (eq. 7); followed by (2) update the appropriate prototype vectors according to the neighbourhood function associated with the BMU (eq. 3). In the case of the standard SOM, iteration would now commence, after potentially updating the learning rate and neighbourhood function. In the case of the ω-SOM, however, an additional step is introduced:

Step 3: Update weights, ω_n, for the currently selected input vector, $\mathbf{x}$ and its associated BMU prototype vector, $\mathbf{m^b}$.

The weights may be updated (Huang *et al*. 2005) according to:

$$\omega_n = \begin{cases} 0, & \text{if } D_n = 0, \\ \left(\sum_i^N \left[\frac{D_n}{D_i} \right]^{\frac{1}{\beta - 1}} \right)^{-1}, & \text{otherwise.} \end{cases} \tag{8}$$

where

$$D_n = \sum_k h_{bk} (x_n - m_{kn})^2 \tag{9}$$

and h_{bk} is the neighbourhood function for the current BMU b and prototype vector k (eq. 3).

We next comment on the role of the parameter β that appears in eqs (7) and (8) and start by examining the parameter D_n. eq. (9) suggests that if data in dimension n are tightly clustered (so that ω_n is expected to be large) then D_n will be small; while more scattered data with a smaller ω_n will correspond to a larger value of D_n. With this in mind, we can examine the relation between ω_n and D_n in eq. (8) and the role played by β. If $0 < \beta < 1$ this would lead to an outcome opposite to that expected; dimensions that show good clustering (small D_n) would be assigned small weights. If $\beta > 1$, a smaller D_n corresponds to larger ω_n and ω_n^β, which is in accord with the idea of feature weighting. If $\beta < 0$, a larger D_n would lead to a larger ω_n, which seems counter to our objectives, however, ω_n^β would in turn be larger again. We remark that $\beta = 0$ is the trivial case when the weights have no impact, according to eq. (7), and the behaviour would replicate the standard map. If $\beta = 1$ and $\omega_n = 1$ for the smallest D_n, only feature n would contribute to the SOM clustering and all other dimensions would be ignored. These observations imply that the closer β is to (but larger than) 1, the stronger the effect of applying weights will be. Features that contain distinct clustering will contribute more to the overall training of the map while those exhibiting less clustering may not contribute at all. In the examples that follow, we consider only cases where $\beta > 1$.

To ensure that training occurs smoothly, updating of the weights is performed using a moving average window. Depending on the input data point used to update the map, it is possible that the ω for a particular feature becomes very large (close to 1) or small (near 0). This has implications for subsequent training steps and can lead to situations in which ω_n does not return to reasonable values (similar to finding in a local minimum). For this reason, the actual ω_n values used in the distance calculation during iteration i are an average of

the ω_n from the previous L iterations, so that

$$\omega_{n,i} = \frac{1}{L} \sum_{l=i-L-1}^{i-1} \omega_{n,l.} \tag{10}$$

Empirically, it is found that L does not need to be particularly large to ensure smooth training and updating of the weights. Values of around 10 or larger were found to be sufficient, although much greater values, on the order of a few hundred, may be preferred for smoother training, especially if data are highly scattered or if there are numerous features.

There already exists a variant of the SOM, also called the Weighted SOM, or WSOM (Sarlin 2012). Although this method shares the name with our work presented here, it differs in how the concept of weighting is used. In our implementation, the weighting is applied to the N different data dimensions but is constant across all M input data vectors, so that different dimensions contribute differently to the clustering procedure. In the variant proposed by Sarlin (2012) the weights are applied to the M data vectors instead so are effectively constant across the N data dimensions (for each input data vector). The main distinction is that different input data vectors make varying contributions, rather than the different data dimensions. There is also a more fundamental difference between the two methods that is worth highlighting. In the implementation we suggest here, the weights assigned to the different dimensions are an emergent output of the method, so that the method provides the user with additional insight into the data structure. By way of contrast, in the version of Sarlin (2012), the user decides a priori the weights to be assigned to the different data vectors and then applies the method. This implies that although the weights of the WSOM approach may help optimize the performance, the method cannot provide additional understanding of the data and may also suffer from possible subjective biases when the weights are decided.

4 APPLICATION TO SYNTHETIC DATA

To test the potential of our ω-SOM algorithm, to classify seismic events based on waveform features, we generate a synthetic data set (Fig. 2) with the following characteristics:

(i) 40 000, 6-D data points.

(ii) Data point values in dimensions 1–5 are distributed according to 10 different normal distributions with means selected randomly between 0 and 1 following a uniform, random distribution.

(iii) Standard deviations of all 10 normal distributions are set to 0.01 for dimension 1 and 0.05 for the other 4 dimensions.

(iv) Data in dimension 6 is distributed according to a uniform distribution.

4.1 Comparison of the standard and weighted SOM variants

We now compare the results of applying the different SOM algorithms to the synthetic data set: the standard SOM (unweighted), and the ω-SOM for different values of the β variable in eq. (7) (β = 1.2, 2 and 5). This is done by training the four SOMs in parallel, with the weight smoothing window length set to 500 iterations, as appropriate.

The four maps are initialized with the same prototype vectors, and the training is done in parallel so that the update on each iteration uses the same input data point for all map variants. Fig. 3 compares the trained maps and associated unified-distance matrices (U-matrix) for the standard SOM and ω-SOMs. The cells of the trained maps are coloured according to the number of input data points for which that cell is the BMU. White or empty cells represent parts of the map for which the cells were not the nearest neurons to any of the input and can be considered as bounding separate, well defined clusters. Warm colours, or high values, therefore represent areas of the map that correspond to high density regions of the input data. The U-matrix is larger than the main map and has cells inserted between all neighbours of the map. The value assigned to these cells is equal to the Euclidean distance between neighbouring SOM neurons in the N-dimensional input data space, $\mathbb{R}^N$. The value of the U-matrix cells corresponding to the SOM cells is equal to the average of the all neighbours (distances to other original SOM cells).

All variants of the map were able to identify and reconstruct the 10 clusters (Fig. 3). This is observable through the separated regions in both the maps (left) and U-matrices (right), where the regions in the main maps are separated by cool colours (or white gaps), while the regions in the U-matrices are separated by the warmer borders. Although all variants of the map create these 10 regions representing the clustered, input data, visually, they appear different in terms of the configuration within the map. As the training was done in parallel, with the same data points being used in the same sequence for all maps, one may have expected that the resultant maps should have the same final configuration. The observed result is due to the fact that the introduction of β changes the calculation of the neighbourhood function, and, subsequently, the re-arrangement of the cells of the map as the updated neuron positions will now differ.

The most notable difference between the maps (Fig. 3) is for the ω-SOM, where $\beta = 1.2$, which exhibits extreme clustering around a few neurons in each of the 10 clusters, with single neurons being considered the BMU for up to 1409 input data points. This is significantly higher than for the other maps, where a maximum count of 150 was observed for $\beta = 2$ and 105 and 106, for $\beta = 5$ and the unweighted maps, respectively. This indicates that although all four algorithm variants were able to identify the clusters, the introduction and value of β impacts the final shape and distribution within the map, with larger values of β leading to behaviour closer to that of the standard SOM. This observation is remarkable considering that the maps were initialized as equal and that the maps were trained using an identical sequence of input data.

4.2 The impact of the parameter, β

We now examine the effect of the different β values on the weights for the different features to further illustrate the impact of this parameter. The evolution of feature weights for the different variants during training (Fig. 4) shows that the larger the value of β, the less fluctuation there is from an equalized weight of 0.17. This applies for the variability during training, as well as the final weights at the end of training, summarised in Table 2.

For high $\beta = 5$, the final weights range from 0.12 to 0.18 while for a lower $\beta = 1.2$, the final weights show a significantly larger range from 0.00 to 0.52. The highest value corresponds to the weight for Feature01 which the map identified as the feature containing the most distinct clustering. Furthermore, for $\beta = 1.2$, the uniformly distributed Feature06 has its weight reduced to zero, indicating that

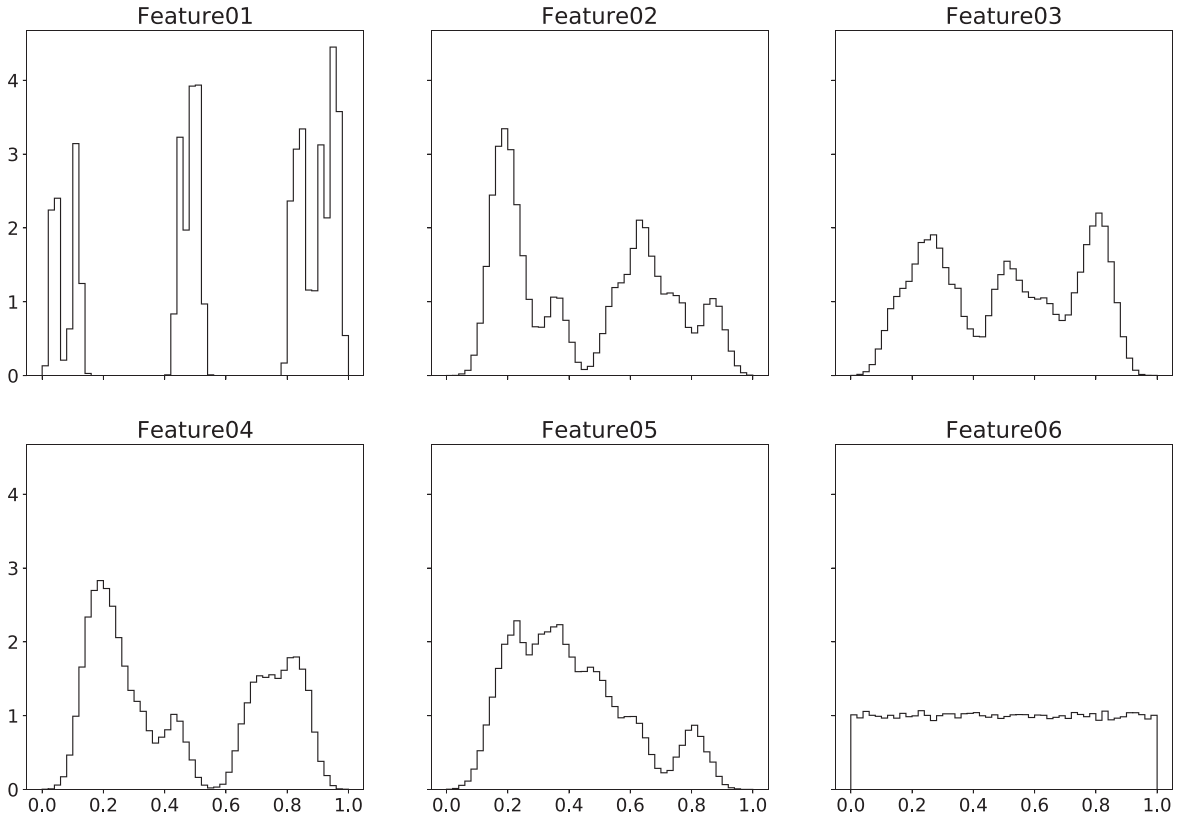

Figure 2. Synthetic data set used to test the ω-SOM algorithm. The distribution of data for the 6 dimensions is shown: Feature01 appears in clear, well defined clusters while Feature06 has a uniform distribution.

when the map is trained, the values of this feature provide no input to the training at all and it could have been excluded from the training at the outset. Features 02, 03 and 04 were also assigned very low, but non-zero values at the end of training, suggesting that they were also identified as relatively unimportant for clustering. In contrast, Feature05 was assigned a relatively large weight (0.37), meaning that when $\beta = 1.2$, the ω-SOM utilized data almost entirely from Feature01 and Feature05, effectively ignoring the others. As such, the β parameter determines the strength with which the weights are applied. Lower values of β amplify the effect of the weight of a feature while if β is increased this amplification and the behaviour of the ω-SOM tends towards that of the standard SOM.

It is worth noting that when $\beta = 0$, eq. (7) reverts to eq. (6) and we retrieve the original, standard SOM. We also remark that as β increases so the behaviour tends towards that of the standard SOM. It would seem rather peculiar that the $\beta \to 0$ and large β behaviours would be the same but this can be ascribed to different equations that control matters in these limits. When $\beta = 0$, $\omega_n^\beta \equiv 1$ within eqs (7) and (8) is redundant. On the other hand, as $\beta \to \infty$ so eq. (8) predicts that $\omega_n \to 1/N$ which reproduces the behaviour of the standard SOM. It should also be remembered that owing to the nature of eq. (8) any small changes in the size of β when this parameter is close to 1 can have quite profound effects on the weights.

4.3 Comparison of component planes

Fig. 5 compares the component planes of Feature01 (strong clustering) and Feature06 (no clustering) for the four map variants. The nature of the distribution of Feature01 for the different maps has a similar character; each one is able to identify and delineate clusters and matches the distribution of clusters in Fig. 3. For Feature06, however, the component plane plots are very different. The standard SOM (unweighted) shows evidence of the cluster separation in the upper left region, with the centre and right regions being without clear cluster separation, and the component plane for $\beta = 1.2$ shows an apparently random distribution of the values with no discernible clusters or trends. The component planes for $\beta = 2$ and $\beta = 5$ show some weak clustering, with boundaries between different clusters generally showing a gradual transition. The distribution of data across the different component planes for the different features is consistent with their assigned input distributions and the influence of the weights. For Feature01, the data in this dimension did contain distinct clustering and all variants of the SOM are able to capture this. Feature06, which contained no clustering, is assigned a weight of 0 when $\beta = 1.2$ and data from this dimension in the clustering procedure is ignored. Subsequently, there is little evidence of the final clusters for that component plane. The standard SOM and $\beta = 2$ and $\beta = 5$ variants still attempt to utilize the data from Feature06, so some trends are visible, although clusters are not explicitly delineated.

The introduction of the ω-SOM and the influence of β are well illustrated using this synthetic data set. By assigning weights to the different dimensions of the input data, the ω-SOM is able to identify which dimensions of the data contribute more or less to the identification of distinct clusters. When β is low, such as the $\beta = 1.2$ case, uniformly distributed data containing no information are assigned a weight of 0. Using such a low value of β may seem appealing,

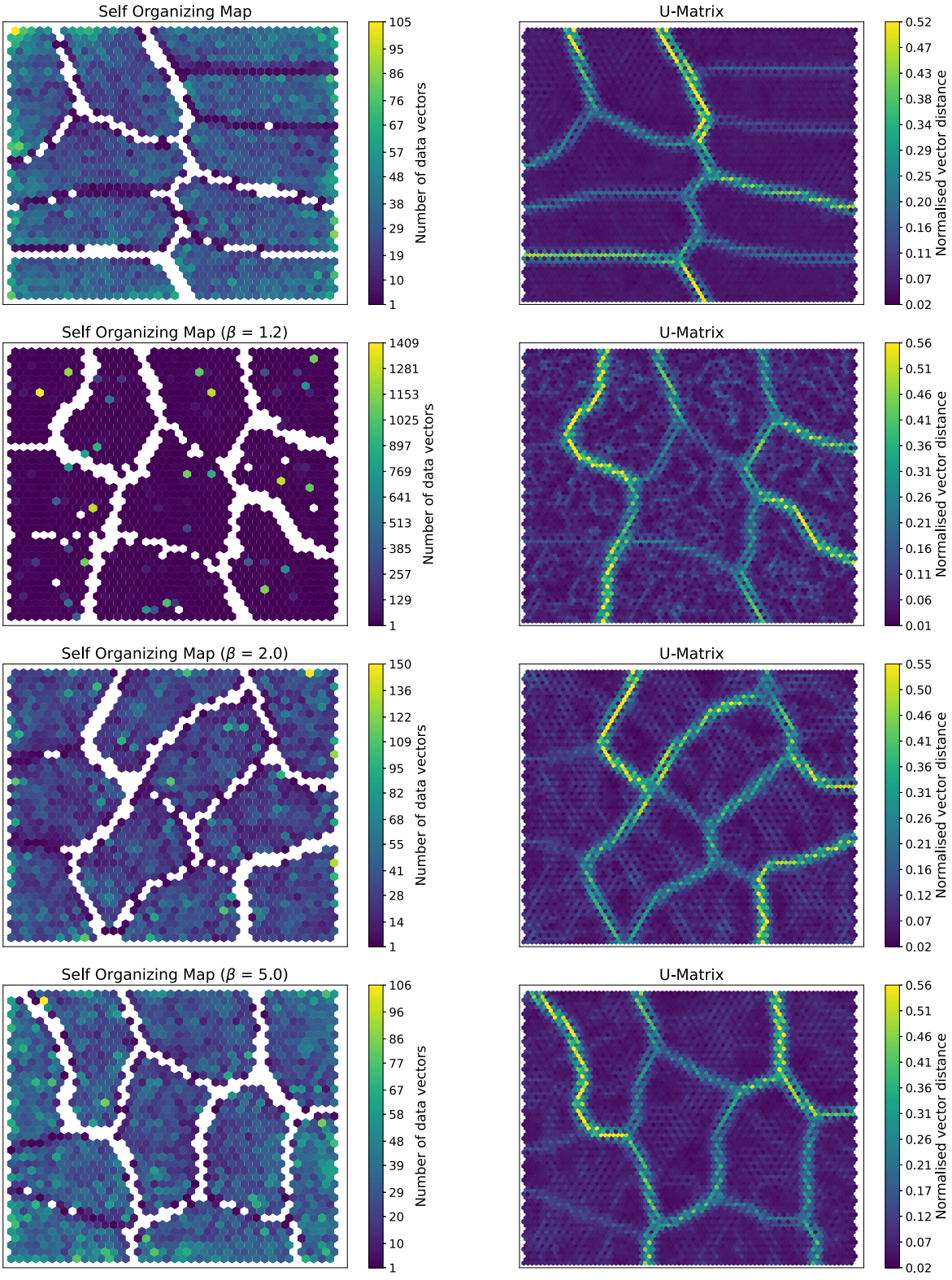

Figure 3. Synthetic data test. Comparison of SOM training on the same data set for the standard SOM (top panel) and ω-SOMs for $\beta = 1.2$, 2 and 5 showing the trained maps (left-hand column) and associated U-matrices (right-hand column).

but for real world data the importance of a single dimension can be overemphasized, which may not be representative of the entire data set. A moderate value for β, perhaps somewhere in the range 2–4, is recommended for use on real data as the training will not be dominated by a single dimension and yet will benefit from the clustered structure of data.

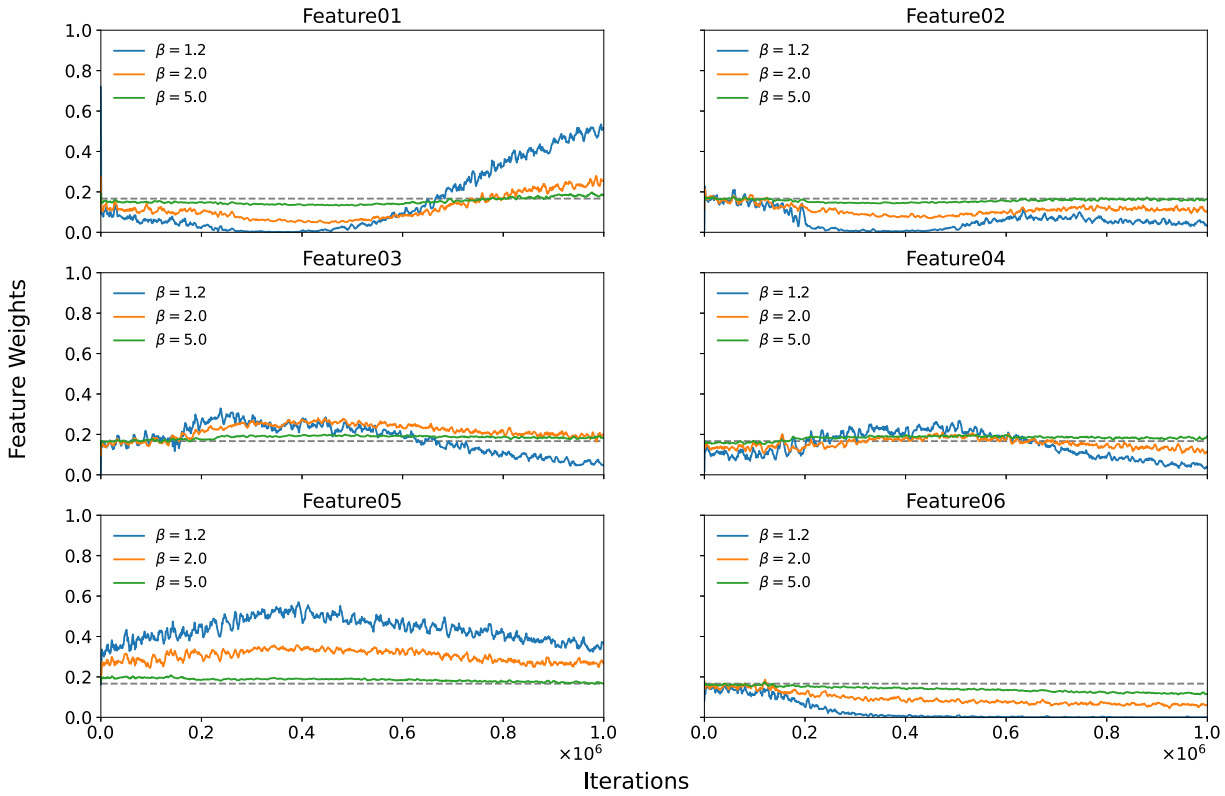

Figure 4. Evolution of the weights for each feature for the three variants of the ω-SOM. Dashed grey line represents the weight value if all features had equal weights. Details of the final weights are listed in Table 2.

Table 2. Synthetic data test weights for the different features (Fig. 2) and β variants at the final iteration of training.

	$\beta = 1.2$	$\beta = 2.0$	$\beta = 5.0$
Feature01	0.52	0.25	0.18
Feature02	0.03	0.10	0.16
Feature03	0.05	0.20	0.18
Feature04	0.04	0.11	0.19
Feature05	0.37	0.27	0.17
Feature06	0.00	0.06	0.12

5 APPLICATION TO REAL DATA

We now apply the ω-SOM algorithm to a real data set of 280 000 seismic records, trained on a 40×40 map. Initially, a 50×50 map was tested, based on the $5\sqrt{K}$ rule-of-thumb. The map dimensions were then reduced slightly, as the results from the smaller map were more appealing with higher density regions. The features used in the analysis are those listed in Table 1 and their normalized distributions are shown in Fig. 1.

5.1 Comparison of the standard and weighted SOM variants

As with the synthetic test, we train a standard SOM and ω-SOM using the same initial map configuration and sequence of records when training. In the case of the ω-SOM, the weights are initialized to have equal distributions and we use $\beta = 4$ and $L = 1000$.

The resultant maps applied to the real data are shown in Fig. 6, for both the standard (top panels) and weighted (bottom panels) variants. Qualitatively, the two maps appear quite similar, with a region of the U-matrix containing high values located in a corner of the map, the lower left for the standard map and upper right for the weighted map. There are also smaller zones of relatively high values on the U-matrix located on the side of the map opposite to the high value region just mentioned. The maps can therefore be regarded as having many characteristics in common, although the relative positions and distributions across the map differ. This means that, even though they were initialized with the same configuration and updated using the same sequence of input vectors, the inclusion of the weighting aspect led to a completely different configuration of the map. Such behaviour was observed in the synthetic case in Fig. 3, where all variants were able to reconstruct the clusters but in different configurations on the final map. This is indicative of the ability of SOMs to preserve the internal topological structure of the data and indicates that the final resultant maps are effectively similar, even if they appear somewhat different.

The maps in Fig. 6 do not show such clear clustering and separation as was observed during the synthetic test. This is expected as the synthetic data set was designed to contain distinct clusters for illustrative purposes. Although the maps of real data show less distinct clustering, they again provide a tool with which to probe the data set, and useful information can be inferred regarding structures contained within the data set. Section 6 includes discussion of how this map can be utilized to realize practical benefit and how feature weights have the potential to give insight into underlying microseismic event processes.

6 INTERPRETATION AND DISCUSSION

6.1 Map validation using independent information

The databases of in-mine seismicity used in this study have the advantage, in the context of method development, that the

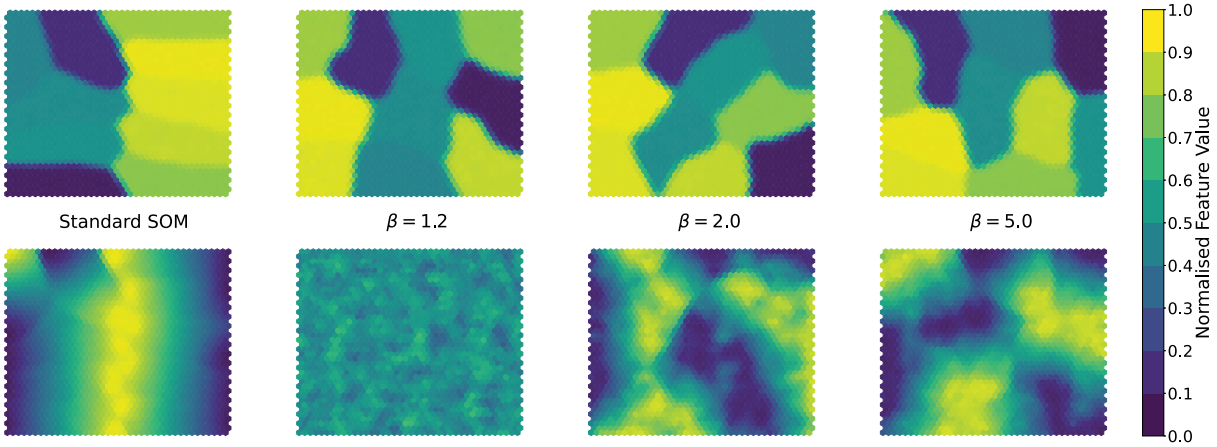

Figure 5. Synthetic data test comparison of the component plane plots for Feature01 (top row) and Feature06 (bottom row) for, from left to right, the standard SOM, and the ω-SOM for $\beta = 1.2$, $\beta = 2$ and $\beta = 5$.

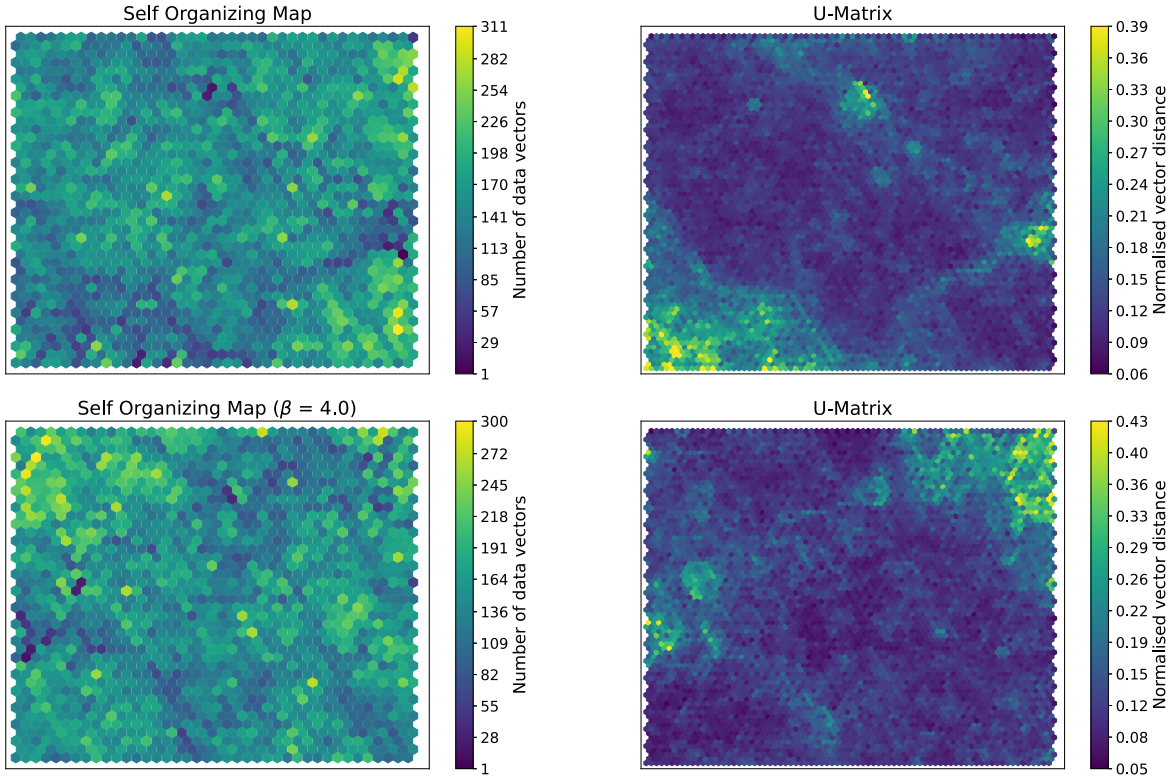

Figure 6. Comparison of SOM results for real data: standard SOM (top row) and ω-SOM with $\beta = 4$ (bottom row).

majority of records have been reviewed and classified, and arrival times picked (when appropriate). The classification and picking is performed by a combination of automated procedures (with review) and trained humans. This information, independent of the SOM applications, allows us to investigate how these already classified events are distributed across the ω-SOM.

For the time period in question, there were 3683 genuine microseismic events out of the total of 283 136 records, constituting a mere 1.3 per cent. As these events have been independently classified, we are able to examine how genuine and discarded events are distributed across the ω-SOM (Fig. 7). Good separation is achieved with genuine events mostly confined to the upper right corner of the map (the highlighted area). This region of the map also shows a much lower density of discarded events, indicating that the map should generally provide a successful tool to separate the events of interest from those to be discarded.

A few isolated contrary cases occur, where events identified as genuine are located in regions of the map otherwise dominated by discarded events. These represent instances of misclassification which is discussed in the next section.

6.2 Limitations

Application of the ω-SOM to the large demonstration data set illustrates how the method is able to separate the majority of events

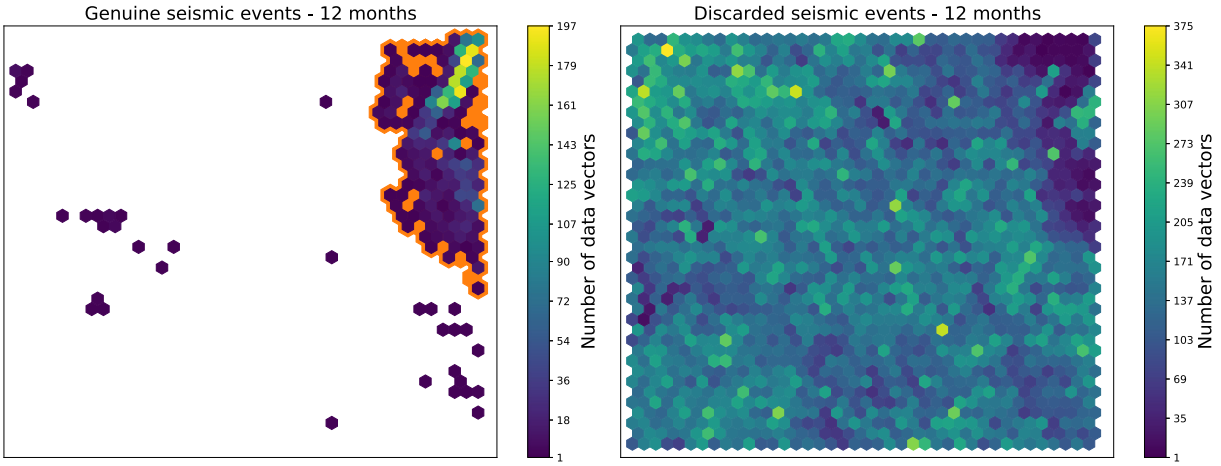

Figure 7. Distribution of genuine seismic events (left-hand panel) and discarded signals (right-hand panel) on the ω-SOM (Fig. 6 bottom). The left-hand plot indicates the area dominated by genuine events (highlighted by the orange domain boundary.

of interest from those that can be discarded, but, the method is not without its limitations. Some situations are encountered in which the SOM, be it either the standard form or a weighted variant, makes an incorrect classification. Fig. 8 demonstrates two such cases where this may happen. In the first example the recorded signal contains two (or more) different sources within the same waveform buffer; in the case shown there is a small genuine event at the start of the buffer but a low frequency noise produced by machinery follows approximately 3 s later, Fig. 8, top). A trained human analyst would identify the small signal at the start of the waveform and process it accordingly. By way of contrast, as the features provided to the SOM are extracted from the entire waveform, the higher amplitude, lower frequency portion may dominate. The resultant features associated with this waveform are therefore more likely to represent the low-frequency portion and hence the event might be classified to be discarded. In this demonstration data set, this would have occurred in about 1.4 per cent of all events classified as genuine. While any loss of such genuine events should ideally be minimized, this loss is not of major concern as small microseismic events such as these fall below the system-wide minimum sensitivity threshold. This means that many are routinely missed, and are not therefore included in hazard assessments and evacuation initiation procedures.

The second example shown in Fig. 8 illustrates a less serious limitation; here waveforms that could be discarded are classified such that they are actually kept. This sample waveform consists of a high frequency, energetic signal with relatively low amplitude that, in many ways, shares characteristics with small, genuine events. Such instances again only constitute a very small proportion of events classified to be kept in our study, and would be unlikely to contaminate the catalogue. At mine sites where many spikes are recorded, they could be identified by the SOM and classified as discarded.

A final limitation of this approach is simply that data are required for the SOM to be trained. Situations may arise when it would be desirable to immediately apply a SOM to a new data set (e.g. consisting of only 1 d of data), but this may be insufficient to adequately train the map. It would be possible to apply a map trained on another data set if the system and environment are similar, such as another similar mine with the same sensors. In this case, the use of a standard SOM over a weighted one would be preferred, as the importance of different features will be unknown and not necessarily the same as for the data set from which the trained SOM originates.

6.3 Practical benefit of a weighted SOM

In hard rock, underground mines, seismic monitoring has become a standard tool used by engineers to monitor and quantify the rockmass response to mining. In some mines, such as our demonstration example, the vast majority of the recorded signals can originate from machinery within the mine and can ultimately be discarded as they contain no valid or useful information for the engineers. Therefore, these in-mine seismic data sets require reliable classification (Dong *et al.* 2016) and processing (Martinsson 2013; Gal *et al.* 2021) to locate seismic events of interest and evaluate their source parameters and mechanisms.

The region of the map highlighted in Fig. 7 is dominated by genuine events. If any event, including one that is newly recorded, is assigned to the highlighted region of the map then that event should be considered to be potentially genuine, and proceed for further processing and analysis, with any events outside this region being discarded. With this rule applied, 93 per cent of the events would be automatically classified for being discarded, allowing resources to be devoted to signals more likely to contain beneficial information.

Another practical outcome of the trained SOM is the ability to identify quickly waveforms with similar characteristics. A user may have a specific event or waveform of interest or concern, with a need to find similar events. This can be achieved by simply extracting those events which are assigned to the same position on the SOM as the event in question. A range of other useful analyses, such as identifying events with certain characteristic waveforms and analysing their spatial or temporal variation, is therefore made feasible.

These results show how a SOM could be applied in real time to identify events of interest, and discard those which are not. Practically, the difference between using a standard and ω-SOM for this task is relatively minor as both achieve similar outcomes. However, a weighted version with a moderately high value of β is recommended, as any preferential clustering in the data will lead to an

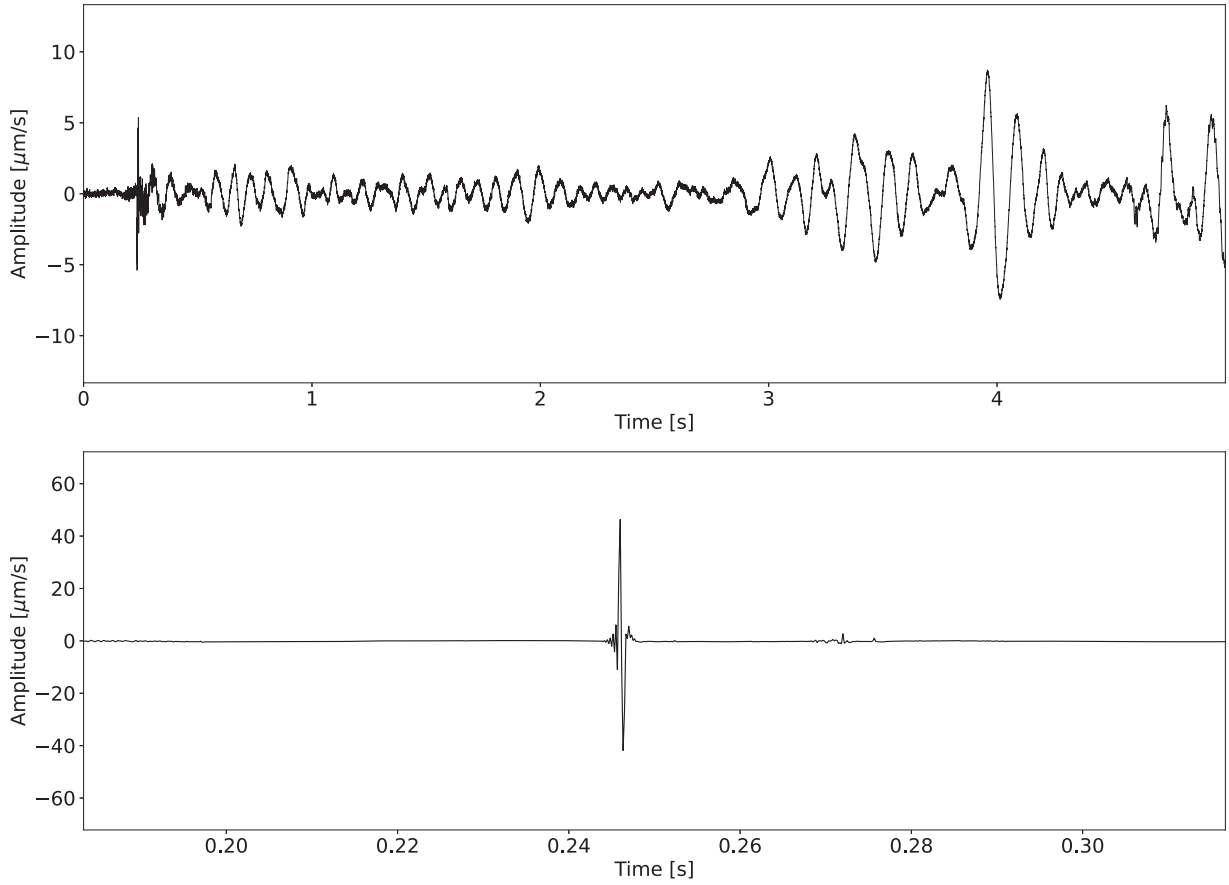

Figure 8. Two example waveforms for which the SOM performed poorly. The first waveform on the left contains a small, high frequency signal at the start of the buffer associated with a genuine seismic event which is followed about three seconds later by low frequency noise from nearby machinery. The waveform on the right contains an electrical spike which could be legitimately discarded but was actually classified to be kept owing to its features being reminiscent of genuine events. In both cases just one (arbitrarily chosen) of the three sensor components is shown.

improvement in the final map structure. The demonstrated test results indicate that applying a SOM could lead to significant time and cost savings as resources can be allocated to more relevant and informative data.

6.4 Insights into evolving source processes

The application of a SOM to large data sets provides practical benefit, irrespective of whether the standard or weighted variant is used. The ω-SOM, however, provides the user with additional investigatory tools with which the probe large data sets. Features with large weights contribute more to data separation and, as such, can be considered as representing characteristic aspects of the generation processes. Features with low weights do not contribute greatly to this discrimination, and, therefore, are not considered to provide a helpful representation of one or more of the generation processes.

The ω-SOM provides additional guidance when one considers the potential evolution of source processes. In order to apply the map to a given sequence of data, an ω-SOM is trained on data taken from the same mine, but from a time period three years previous (Period A). The length of Period A was selected so that the total number of records was similar to the given sequence (Period B, our main study), ensuring that the two maps had comparable data density. The parameters such as size of the map and β are kept the same for the two sets of data. The feature values were normalized according to the minimum and maximum values of both periods combined so that they are directly comparable.

During Period A, mine processes involved fracturing the rock in the target area for ore extraction. In contrast, during Period B, this fractured region is already well established. As the mining was at different stages, both the overall state of the rockmass, and the mining activities being conducted, are likely to differ between the two periods. Period A would likely include more frequent and stronger blasting compared to that in Period B. Moreover the process of initiating major fracturing in the rockmass during Period A could also lead to a relatively high proportion of small, energetic microseismic events associated with the failure of the intact rockmass. During Period B these microseismic events will still occur, but, as the region of broken, mobilized material would be significantly larger, we may expect a larger proportion of low-frequency sources associated with the movement of this material. Furthermore, during Period B, the larger volume of broken material leads to significant changes in the the stress field surrounding the mine. This may initiate the activation of geological structures that were previously quiet, owing to the induced stress changes.

We now focus on the weights assigned to different features and their relative rank, rather than the resultant maps. The final weights after training the ω-SOMs on the two periods are compared (Fig. 9, top panel) and are quite similar in many cases. The rank of each

Figure 9. Top panel: histogram of final weights of the different features for the two time periods. Horizontal line indicates the average, equalized value. Bottom left-hand panel: change in the rank of features between the two time periods, with rank 1 being the largest weight feature. Text annotations refer to feature ID as listed in Table 1. Bottom right-hand panel: comparison of normalized feature distributions between Period A (earlier, blue) and Period B (later, orange) with feature IDs the top corner. Feature 2 is excluded for illustrative purposes.

feature in the two periods illustrates that, similar to the values of the weights, the majority of features experienced little change in their respective rank (Fig. 9, lower left-hand panel). Comparing the distribution of observed values for each feature between the two periods shows that most underwent minor changes (Fig. 9, lower right-hand panel, e.g. Feature 08), although some exhibited a more dramatic shift (e.g. Feature 19) and a few others experienced almost no change at all (e.g. Feature 11). The observed distribution of values associated with Feature 01 changed considerably between the two time periods although the relative importance remained fairly high. This indicates that the source generation processes had evolved but the feature remained useful for separating the data, as would be the case if an existing source generation process were to persist but change in some way.

Changes (or lack thereof) to distributions of feature values are now considered in conjunction with corresponding changes in rank. One may expect that if the observed values of a feature altered markedly, so might the relative importance of that feature and that, similarly, little change in the observed values would be accompanied by no more than perhaps a small adjustment in the relative importance. Feature 03 retained its high rank (1 and 2) over the two periods and the distribution of observed values also remained fairly consistent, an important characteristic of the data in regards to separation of the source processes. Feature 19 presents the opposite, but also foreseeable, behaviour; the feature shows both a dramatic change in the observed values accompanied by a major change in its importance. Other scenarios are perhaps less predictable. For instance, Feature 01 demonstrated a considerable

change in the distribution of observed values but remained of similar and fairly high importance, at rank 7 and 8 in Periods A and B respectively. On the other hand Feature 06 possessed almost identical distributions in the observed values between the two periods but suffered a relatively large change in rank, dropping from 12 to 6.

In cases where both the rank and observed values change considerably, the interpretation is that some characteristic of a source process has evolved, together with the importance of that characteristic for data separation. This could occur if existing source processes were to cease or new ones were to appear (as with Feature 19). In Period A, the source generation processes were such that Feature 19 was found to play an important role in separating the data. However, during Period B, the source processes had changed, as evidenced in the observed distribution of Feature 19 values, and in such a way so that this feature was no longer so pivotal for the clustering. With regard to the evolving source processes, we remark that the interpretation of the changes in the observed distributions and weights in the context of changes to a single source process should be treated with some caution as this ignores the multivariate nature of unsupervised machine learning techniques such as the SOM.

The observed variations in the feature importance (relative weights) and feature value distributions in Fig. 9 implies that there were changes to the source generation processes between Periods A and B. We demonstrate this further by examining some waveforms in more detail. Fig. 10 considers two genuine microseismic events, one taken from each time period. The feature distributions on the right correspond to the genuine events in the respective periods only. For many features, the distributions of the observed genuine events are very similar. Both waveforms shown are from $m_W = 0.0$ events and are recorded at fairly similar hypocentral distances (based on the S-P separation). The vertical coloured bars in the distributions illustrate the feature values associated with the waveforms. Both events are of high amplitude and of relatively short duration, as captured by Features 01 (logarithm of maximum amplitude) and 19 (T90 duration), respectively. The waveform from Period B is seen to contain some higher frequency content, which is also captured by the values of Feature 13. These three examples are cases in which the expected feature value can be roughly estimated based on a cursory visual inspection of the waveform. For many other features, such as Feature 07, this visual technique is infeasible. The values of this feature were extremely similar for the two waveforms shown but they occupy different positions within their relative distributions; the Period A event is near the middle of the distribution, while the Period B event falls on the lower range. The feature distributions for the genuine events are generally fairly similar, but Feature 04 is the clear exception to this. The relative consistency of the distributions of some Features such as 01, 03, 13 and 19 indicate that the changes are likely not due to variations in the source–sensor distributions or distances, but actually represent changes to some aspects of the source generation processes. If the sources were suddenly much further from or closer to the sensors, for example, one could expect a dramatic change in these features; at larger distances amplitudes would be lower and higher frequencies would have been attenuated out.

The fact that minor changes in the distributions of some features are observed indicates that there are more subtle variations within the different seismograms. For example, Features 08, 09 and 10 count the proportion of samples within a seismogram that are above 0.75, 0.50 and 0.25 of the maximum amplitude, respectively. For the seismograms shown (which were for events of the same magnitude, recorded at similar distances), these three features are all higher for the Period A event than the Period B event. This can be deduced from the waveform, as the Period B event contains clear, impulsive arrivals with relatively little scattering, and few other samples that have relatively high amplitude. Conversely, event A has periods directly following the *P*- and *S*-wave arrivals during which the amplitude remains fairly high and subsequently records larger values for these three features. In this way, these three features may be considered as proxies of the amount of scattering within a waveform. A comparison of the distributions of these features shows that, overall, period B events report lower values for these features. The Period A event is situated in the centre of the distributions of these three features and can therefore be considered quite representative. The Period B event, while slightly on the lower end of values according to the Period B distribution, would be on the much lower end of the distribution for Period A. This is interpreted that the events in period B exhibit less scattering in the waveforms than those in Period A. Considering the distribution of observed sources in these periods, the Period A example was taken from an event associated with fracturing of the rockmass on the perimeter of the region of broken material. The events here are fairly energetic but, due to their proximity to the broken material, can exhibit high levels of scattering immediately following the arrivals. On the other hand, the Period B example waveform is associated with slip on a geological structure which was not active during Period A due to the induced stresses at that time being insufficient to initiate sudden slip. As the event originated on this geological weakness remote from the excavations and regions of broken material, the waveforms are much cleaner and exhibit very little scattering.

Fig. 11 repeats this exercise but for two events that were classified as discarded. The exact source of the events is unclear, although the Period A event is likely due to some mechanical noise or machinery, while the Period B event is perhaps associated with the movement of loose, broken rock. In this case, we see that the observed distributions (which are for the discarded events only) are much closer to the overall distributions (Fig. 9), as these discarded events constitute the majority of recorded signals. The feature distributions of these discarded events generally contain larger differences than those for the genuine events, indicating the there is more distinct variability in the sources responsible, while the genuine events contain more subtle changes. The most apparent case for this would be an identifiable source of a noise (e.g. a specific piece of equipment in some fixed location) becoming active, or ceasing. This would result in dramatic changes in the feature distributions, as observed, leading to changes in feature weights evaluated by the ω-SOM as the different features contain relatively more or less useful information with the appearance or disappearance of specific sources.

These observations have implications for the choice of features when performing new or updated studies within a region or domain. Since the nature of the sources can vary, it should not be assumed that that a SOM trained on one period will necessarily be applicable to another. The weighted variant of a SOM therefore has additional utility for guiding expert decision making in understanding whether a shared model is appropriate or if the characteristics of two particular data sets are moderately similar, but include an evolution in source processes.

The ω-SOM has been shown to identify which features contribute more and which less to the clustering process, allowing for the poorly performing features to be removed which can lead to

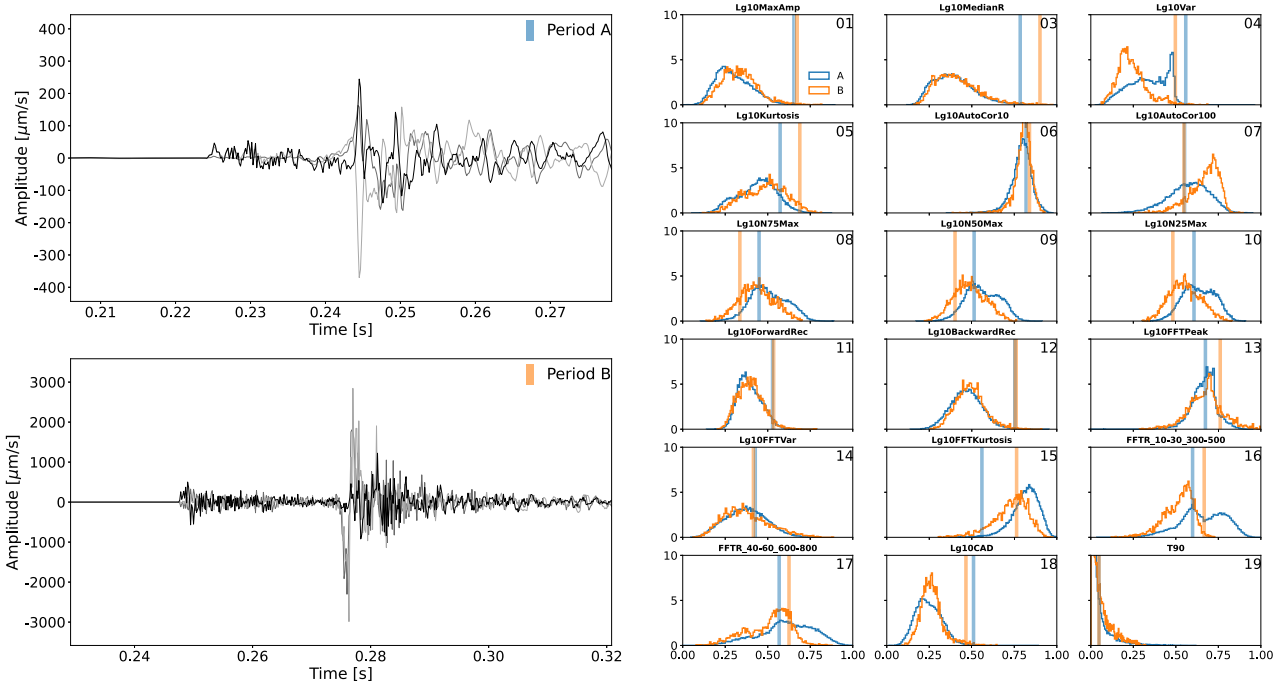

Figure 10. Example waveforms from two $m_W = 0.0$ genuine seismic events, one from Period A (top, blue) and one from Period B (bottom, orange); waveform component description provided in caption to Fig. 1, the corresponding ω-SOM for Period B is shown in Fig. 7. The associated features are plotted as vertical bars along the distributions for genuine events from the two times periods. Feature values are normalized using the complete set of all events from both periods.

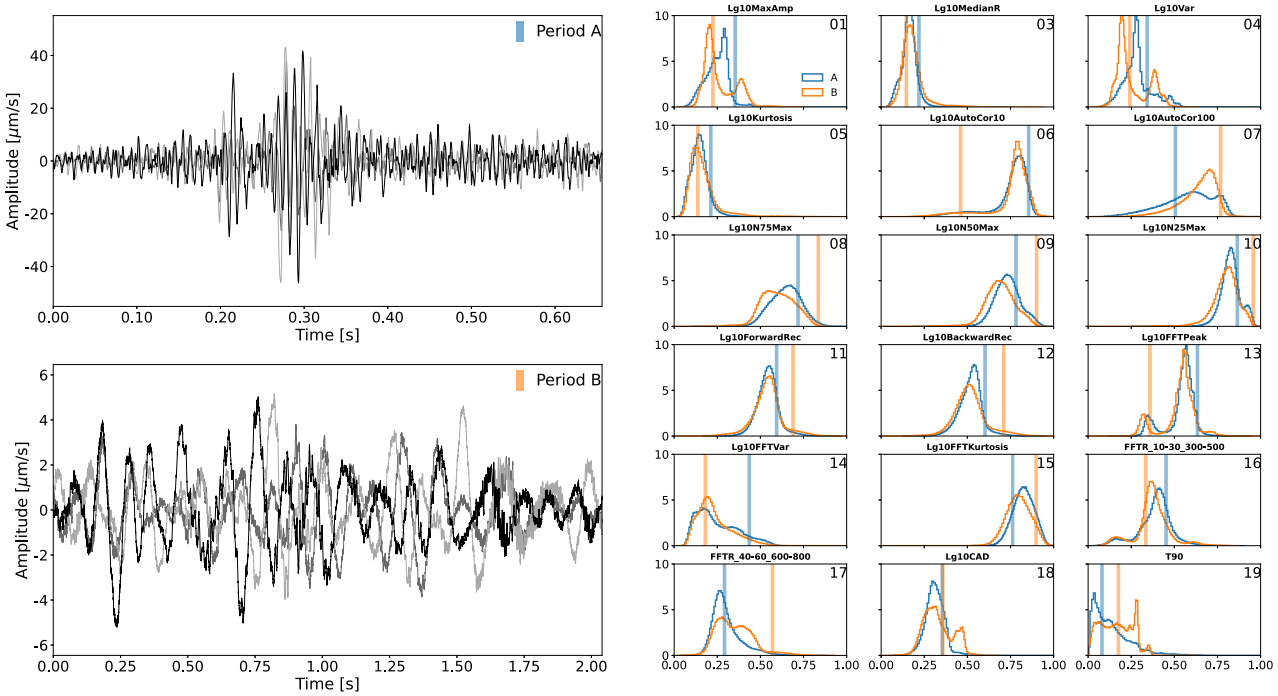

Figure 11. Example waveforms from two discarded seismic events, one from Period A (top, blue) and one from Period B (bottom, orange); waveform component description provided in caption to Fig. 1, the corresponding ω-SOM for Period B is shown in Fig. 7. The associated features are plotted as vertical bars along the distributions for genuine events from the two time periods. Feature values are normalized using all records, the same as in Fig. 10.

potential computational cost savings. However, keeping all features can be beneficial as the importance of features may change as new data become available, as shown in Fig. 9. The question whether features with little contribution should be kept or not depends on the use case. If one has a very large data set where the sources are believed to be fairly constant, it may be preferable to train an ω-SOM on a small, representative sample of it. Features that perform poorly can be identified and removed before retraining on the entire

data set. In other cases, where one believes there may be variations in the source processes, it is preferable to keep all features. Generally speaking though, training of the SOM is done very quickly, so unless the data set or number of features are incredibly large, seeking computational time improvements by reducing the number of features is likely not required.

7 CONCLUSIONS

We have applied the weighted variant of a SOM algorithm to large seismic data sets. Our case studies use signals recorded in an underground mine that arise from a diverse set of microseismic and anthropogenic sources. The previous appraisal of the data by human analysts allows for verification of the unsupervised classification. The training of the SOM utilizes waveform features that are quick to extract from, or calculate using, the recorded waveforms, allowing it to be applied quickly and in real time, on data sets of significant size.

Results from the standard SOM are compared with the weighted SOM variant. When data features show reasonable separation and the ability to be clustered, the weighted SOM leads to improved separation of clusters as it assigns less weight to features that reduce clustering. Identification of the features with the lowest weights allows for these redundant features to be removed, as appropriate, in future classification. This is of practical benefit in cases where a very large number of features are available but it is found that the data can be successfully clustered using a subset of these features. Once the SOM is trained, a user can utilize the map to quickly find subsequent signals with high similarity to previously identified clusters, or unique signals of which only a few are present. In the demonstration example, the large number of anthropogenic signals can be discarded, enabling focus on the genuine microseismic events. This provides expert users with a practical and unbiased tool to interrogate a data set, allowing them to gain a greater understanding of the content and underlying structure.

The weighted variant of the SOM can be used for deeper investigation into the features characterizing a data set, providing insights into the underlying event generation processes. Features with large weights contribute more to cluster separation and can be considered more representative of the source generation processes. This assists with the identification and characterization of specific source processes occurring within a complex domain.

Applying an ω-SOM to two data sets taken from the same domain, but from different time periods, provides a means of identifying the evolution of the processes occurring within that domain. Significant changes in the importance of features and distribution of feature values suggests fundamental changes in the underlying source mechanisms such as specific aspects of processes changing, new processes becoming established, or processes ceasing.

Unsupervised techniques such as the ω-SOM are powerful methods that capture the complex and multidimensional nature of such data sets, in which relations between different processes and features are not evident to the human eye. Recent studies utilising machine learning in seismology often focus on advanced, supervised techniques. These are complex and often somewhat opaque in their behaviour and can sometimes seem to act like a 'black box'. We demonstrate that unsupervised techniques still deliver significant value in the study of large and complex data sets, by not only achieving practical outcomes such as event classification, but also provide the user with the tools to interrogate the data set. Furthermore, this is done in such a way that can be easily visualized and interpreted using the resultant maps, U-matrices and component planes. Large and intricate data sets are presented in a manner that raises questions that the expert may not have considered, ultimately revealing properties of the data that might otherwise have remain hidden.

ACKNOWLEDGMENTS

This study has been completed as part of a PhD at the University of Tasmania.

The referees and editor are thanked for their insightful suggestions that helped improve the manuscript.

DATA AVAILABILITY

The data used in this study are privately owned and not openly available for distribution. All code used for the SOM algorithm was developed for this study and written in python. Reasonable requests made to the corresponding author for examples of data or code are welcome.

REFERENCES

Ahmad, A.U. & Starkey, A., 2018. Application of feature selection methods for automated clustering analysis: a review on synthetic datasets, *Neural. Comput. Appl.,* **29**(7), 317–328.

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology,* University Science Books.

Chaudhary, V., Bhatia, R.S. & Ahlawat, A.K., 2014. A novel Self-Organizing Map (SOM) learning algorithm with nearest and farthest neurons, *Alexand. Eng. J.,* **53**(4), 827–831.

de Bodt, E. & Cottrell, M., 2000. Bootstrapping self-organising maps to assess the statistical significance of local proximity, in *Proceedings of the ESANN"2000'" 8th European Symposium on Artificial Neural Networks,* pp. 245–254. D-Facto, Brussels, Belgium

Dong, L., Wesseloo, J., Potvin, Y. & Li, X., 2016. Discrimination of mine seismic events and blasts using the fisher classifier, naive Bayesian classifier and logistic regression, *Rock Mech. Rock Eng.,* **49**(1), 183–211.

Ekström, G., Nettles, M. & Dziewoński, A.M., 2012. The global CMT project 2004-2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.,* **200-201,** 1–9.

Esposito, A.M., Giudicepietro, F., D'Auria, L., Scarpetta, S., Martini, M.G., Coltelli, M. & Marinaro, M., 2008. Unsupervised neural analysis of very-long-period events at Stromboli Volcano using the elf-organizing maps, *Bull. seism. Soc. Am.,* **98**(5), 2449–2459.

Fort, J.-C., Letremy, P., Cottrell, M., Elie, I. & Nancy, U., 2002. Advantages and drawbacks of the Batch Kohonen algorithm, in *Proceedings of the European Symposium on Artificial Neural Networks ESANN,* pp. 223–230.

Gal, M., Lotter, E., Olivier, G., Green, M., Meyer, S., Dales, P. & Reading, A.M., 2021. CCLoc—an improved interferometric seismic event location algorithm applied to induced seismicity, *Seismol. Res. Lett.,* **92**(6), 3492–3503.

Guérif, S. & Bennani, Y., 2007. Dimensionality reduction through unsupervised features selection, in *Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN 2007),* August 2007, Thessaloniki, Hellas, Greece. ⟨hal-00352375⟩.

Guérif, S., Bennani, Y. & Janvier, É., 2005. μ-SOM: weighting features during clustering, in *Proceedings of the WSOM 2005 - 5th Workshop on Self-Organizing Maps,* Paris, France, September 5–8, 2005, pp. 397–404.

Huang, J.Z., Ng, M.K., Rong, H. & Li, Z., 2005. Automated variable weighting in k-means type clustering, *IEEE Trans. Pattern Anal. Mach. Intellig.,* **27**(5), 657–668.

Ida, Y. & Ishida, M., 2022. Analysis of seismic activity using self-organizing map: implications for earthquake prediction, *Pure appl. Geophys.,* **179**(1), 1–9.

Köhler, A. & Ohrnberger, M., 2008. Unsupervised feature selection for pattern search in seismic time series, in *JMLR: Workshop and Conference Proceedings,* 15 September 2008, Antwerp, Belgium, Vol. **4,** pp. 106–121.

Köhler, A., Ohrnberger, M. & Scherbaum, F., 2010. Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps, *Geophys. J. Int.,* **182**(3), 1619–1630.

Kohonen, T., 1990. The Self-Organizing Map, *Proc. IEEE,* **78**(9), 1464–1480.

Kohonen, T., 2001. *Self-Organizing Maps,* Springer.

Kohonen, T., 2013. Essentials of the self-organizing map, *Neural Networks,* **37,** 52–65.

Kortström, J., Uski, M. & Tiira, T., 2016. Automatic classification of seismic events within a regional seismograph network, *Comput. Geosci.,* **87,** 22–30.

Kriegerowski, M., Petersen, G.M., Vasyura-Bathke, H. & Ohrnberger, M., 2019. A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms, *Seismol. Res. Lett.,* **90**(2A), 510–516.

Langer, H., Falsaperla, S., Powell, T. & Thompson, G., 2006. Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat, *J. Volc. Geotherm. Res.,* **153**(1–2), 1–10.

Marsland, S., 2015. *Machine Learning: An Algorithmic Perspective,* CRC Press.

Martinsson, J., 2013. Robust Bayesian hypocentre and uncertainty region estimation: the effect of heavy-tailed distributions and prior information in cases with poor, inconsistent and insufficient arrival times, *Geophys. J. Int.,* **192**(3), 1156–1178.

Mendecki, A.J., 1997. *Seismic Monitoring in Mines,* Chapman and Hall.

Mendecki, A.J., 2019. Simple GMPE for underground mines, *Acta Geophys.,* **67**(3), 837–847.

Provost, F., Hibert, C. & Malet, J.P., 2017. Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier, *Geophys. Res. Lett.,* **44**(1), 113–120.

Reynen, A. & Audet, P., 2017. Supervised machine learning on a network scale: application to seismic event classification and detection, *Geophys. J. Int.,* **210**(3), 1394–1409.

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J. & Johnson, P.A., 2017. Machine learning predicts laboratory earthquakes, *Geophys. Res. Lett.,* **44**(18), 9276–9282.

Sarlin, P., 2012. A weighted SOM for classifying data with instance-varying importance, in *Proceedings of the 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012,* 10 December 2012, Brussels, Belgium, pp. 187–193.

Tibi, R., Linville, L., Young, C. & Brogan, R., 2019. Classification of local seismic events in the Utah region: a comparison of amplitude ratio methods with a spectrogram-based machine learning approach, *Bull. seism. Soc. Am.,* **109**(6), 2532–2544.

Zhu, W. & Beroza, G.C., 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216**(1), 261–273.

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z. & Fang, L., 2019. Deep learning for seismic phase detection and picking in the aftershock zone of 2008 Mw7.9 Wenchuan Earthquake, *Phys. Earth planet. Inter.,* **293**(April), 106261 , doi:10.1016/j.pepi.2019.05.004.