Using clustering algorithms to assist short-term seismic hazard analysis in deep South African mines

DB Rebuli *Institute of Mine Seismology, South Africa*

SJ Kohler *Institute of Mine Seismology, South Africa*

Abstract

Short-term seismic hazard based on the stability analysis methods of the 1990s is part of the deep mine culture of South African mines. Temporal precursory patterns have shown some level of success not only in deep gold mines but also in shallower mines in Australia.

In complex geological and mining environments precursory patterns in temporal trends can be lost due to multiple processes mixing and diluting these trends. It is not always possible to separate these mixed processes by refining the traditional spatial filters, making analysis of these regions difficult. This paper proposes the use of an unsupervised learning algorithm to categorise seismic data with the intention of separating different seismogenic processes. This algorithm should not only take spatial and temporal parameters into consideration but should include parameters related to seismic sources, e.g. energy index, E_S/E_P ratio, etc.

In this investigation, an implementation of the DBSCAN clustering algorithm to categorise seismic data was used. This density based algorithm requires only two parameters: the maximum search distance, ρ , and the minimum number of neighbours required to form a cluster, n_{min}. The advantages of DBSCAN include its ability to create any number of clusters to suit the data set as well as its ability to identify noise and outliers. DBSCAN can struggle to cluster data sets where there are large differences in densities, since there may not be an appropriate choice of ρ and n_{min} for all clusters.

We investigate a few cases where multi-dimensional clustering indicates precursory trends. More specifically examples of density clustering of spatial-and-temporal and spatial-and-source clustering are discussed.

1   Introduction

Short-term hazard assessment has been part of the South African deep mining industry for many years (Brink & Mountfort 1984, p. 317). This has typically involved experienced mine seismologists analysing the seismic trends in multiple regions of a mine on a daily basis. Ideally, the influence of each seismic event should be investigated as it is recorded but for practical and economic reasons, this is generally not possible. This research has focused on ways to automatically determine the seismic hazard of each production area of a mine multiple times a day. Previous studies have focused on the temporal trends of a number of seismic parameters currently used for short-term hazard assessment (van Aswegen 2005, p. 437; Mendecki 1997; Rebuli & van Aswegen 2013, p. 485). Figure 1 is a classic temporal trend showing decreasing energy index (EI) and accelerated cumulative apparent volume (CAV) before a large event (Hills et al. 2013, p. 485).
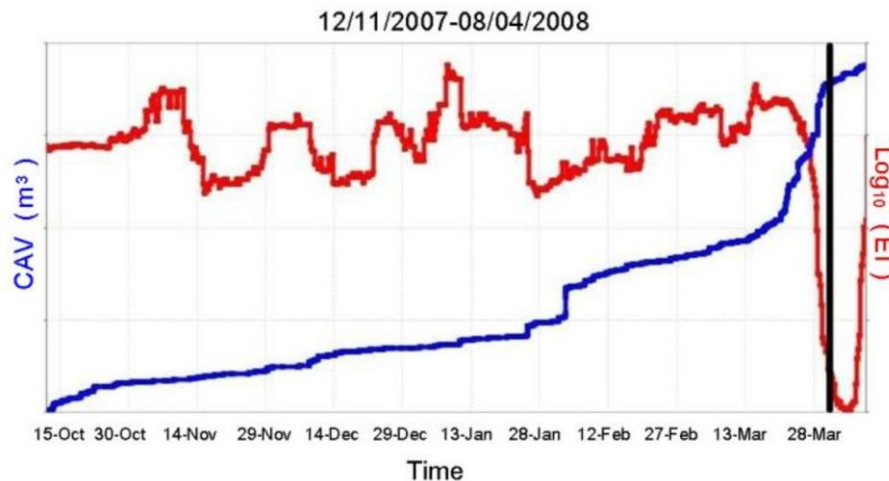
Figure 1 Time history of the period leading up to a large seismic event. The decrease in EI and accelerated CAV are classic instability indicators (Hills et al. 2013, p. 485)

These trends are not always observed before a large event and it is speculated that this may be due to competing processes blurring the temporal trends. Spatial filtering is generally used to try and eliminate the effect but this is not always appropriate. There are other techniques which can be used to extract information from the data. Clustering is one of these techniques and is a class of unsupervised learning algorithms which attempts to group similar seismic events together. There are many different types of clustering algorithms, each with their own strengths and weaknesses. This research focused on clustering algorithms that use a weighted Euclidean metric to partition the data into hard clusters, i.e. a data element can only belong to a single cluster. In particular, the single-link clustering algorithm (SCA) and the density clustering algorithm (DCA) were used to analyse seismic data.

2 Clustering method

2.1 Single-link clustering algorithm

In general, the clustering algorithms would cluster data elements which would be a R^N representation of a seismic event. Each component of the data element would represent a parameter of a seismic event. Single link clustering (SCA) uses a weighted Euclidean metric to determine the distance, ρ, between any two data elements in the parameter space. The distance is defined such that data elements which are similar would have shorter distances than data elements which are dissimilar. In SCA, each data element starts off in its own cluster. The elements are then compared to each other and if ρ is less than the maximum distance parameter, ρ_{max} then the elements are considered to be similar and the clusters containing them are merged. The clusters that remain after all the elements have been compared to each other form the partitioning of the data elements.

This particular algorithm can produce an effect known as chaining. An example of chaining can be seen in Figure 2, where the two clusters A and B are in close proximity to each other. It is possible to incorrectly merge these two clusters by introducing a few data elements between the two clusters. In the context of a mining environment, each cluster should be associated with a single seismogenic region for it to be meaningful when analysing the seismic hazard. The chaining effect of SCA would result in large clusters which are associated with many seismogenic regions making analysis of the seismic hazard difficult. An example of this can be seen in Figure 3 where clusters of seismic events are visible at the panels but the many scattered events would link all the clusters together forming a single large cluster covering multiple panels. A modification to the SCA would help reduce the likelihood of the chaining effect occurring. This new clustering algorithm is known as the DCA.
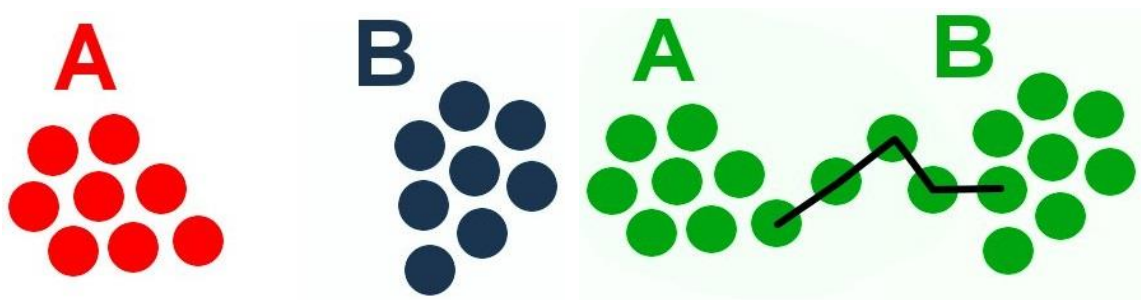
Figure 2   Example of the chaining effect produced by the single-link clustering algorithm; (a) original clusters;
(b) merging of the clusters due to the chaining effect resulting from a couple of stray events between the clusters
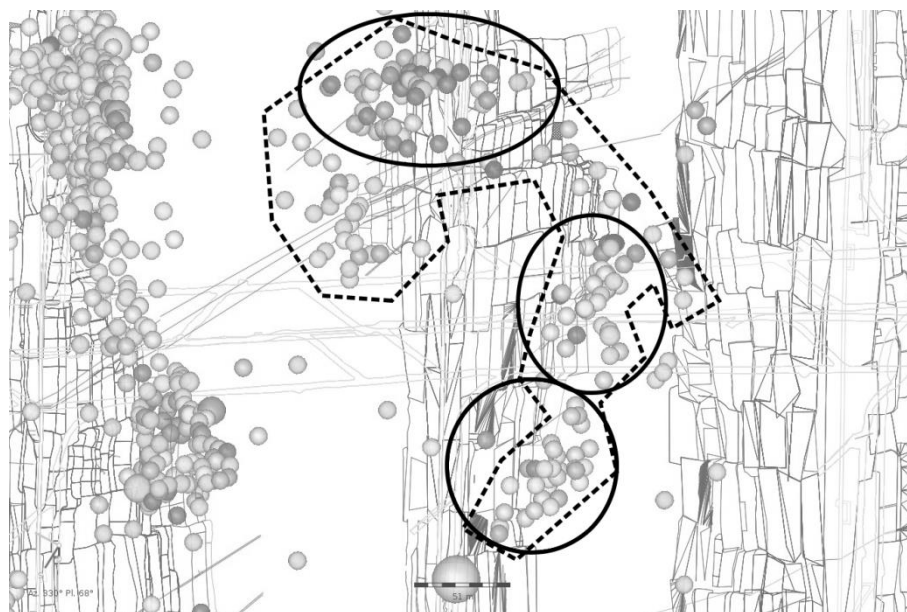

Figure 3   Example of the chaining effect in the mining environment. Three distinct clusters can be observed around the panels (outlined ellipsoids), however, the scattered events located between the clusters could cause SCA to merge all the clusters into one large cluster (indicated by dotted freehand shape)

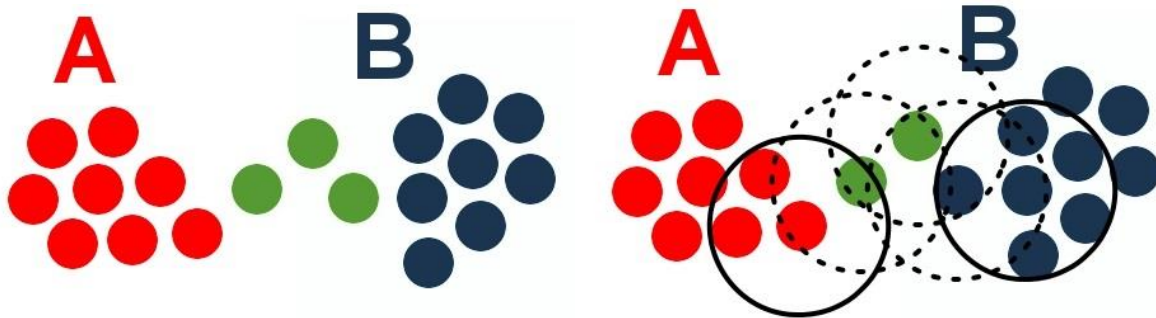## 2.2   Density clustering algorithm

This density clustering algorithm is an implementation of the DBSCAN clustering algorithm (Ester et al. 1996). It improves upon the SCA by adding a minimum number of neighbours parameter, n_{min} . Two data elements are said to be neighbours if $\rho < \rho_{max}$. The parameters, n_{min} and ρ_{max}, are used to classify data elements as one of the following:

- Core element: a data element which has at least n_{min} neighbours.

- Boundary element: a data element which has at least one core element as a neighbour but has fewer than n_{min} neighbours.

- Noise element: a data element which has fewer than n_{min} neighbours with none of them being core elements.

The algorithm works by finding an unvisited core element and creating a cluster containing the core element and its neighbours. The algorithm then recursively goes through all the unvisited core elements in

the cluster and adds their neighbours to the cluster. This process repeats itself until all the core points have been visited. This algorithm would return a set of clusters and list of all noise elements.

The combination of n_{min} and ρ_{max} means that two clusters would only merge if they shared core elements. This reduces the possibility of chaining as chaining is mostly due to data elements on the fringes of the clusters linking together. For this reason, the DCA was chosen over the SCA as the preferred clustering algorithm for this research.

Figure 4   **Example of density clustering algorithm overcoming the chaining effect. The original clusters A and B are shown with the additional elements between them (a); the resulting clusters A and B are shown (b) with**
B growing by a single new element. The black circles indicate ρ_{max}, and n_{min} was set as 5

2.3   Seismic events

There are several issues which should be taken into consideration when clustering seismic events:

1. How many event parameters should be used to cluster the data.

2. How a seismic event should be mapped to a data element.

3. How the different units and orders of magnitude should be handled in the weighted Euclidean metric.

In general, one would need to create a function which would map an event parameter to a real number. Typically, the parameters that one would be interested in, are already quantified (e.g. x coordinate, time, energy, moment, etc.) and thus the mapping would be trivial. Each event parameter would then be represented by one component in the data element. There is no limit to the number of event parameters that can be used in the clustering algorithm but it should be noted that increasing the number of parameters would increase the dimension of real space thereby increasing the distance between the data elements. Using the minimum number of parameters needed to describe the clustering problem would generally produce better results.

A value range standardisation procedure is used on each of the parameters in order to handle the units and the order of magnitude of the parameters. The standardisation procedure (Milligan & Cooper 1988) is as follows: $u_i = \frac{u_i' - u_{imin}}{u_{imax} - u_{imin}}$ where $u_i \in [0,1]$ is the ith component of the data element u, u_i' is the value of ith parameter, $u_{imin/max}$ is the minimum/maximum value of the ith parameter.

The resulting data element, u, which represents the N parameters of the seismic event, can be described as $u \in R^N$ and $u_i \in [0,1]$.

3   Results and discussion

## 3.1   Spatial-temporal clustering

Figure 5 show a section of a mine together with production points and all the seismic events recorded during a 24 hour period. The purpose of this investigation is to generate a set of clusters and classify them according to given production data. For a cluster to be associated with production it would need to be both spatially close to a production point and around blasting time. It is interesting to note that while there are many production points there are only a few event clusters. One can identify six clusters by eye (marked A-F).

(a)   (b)
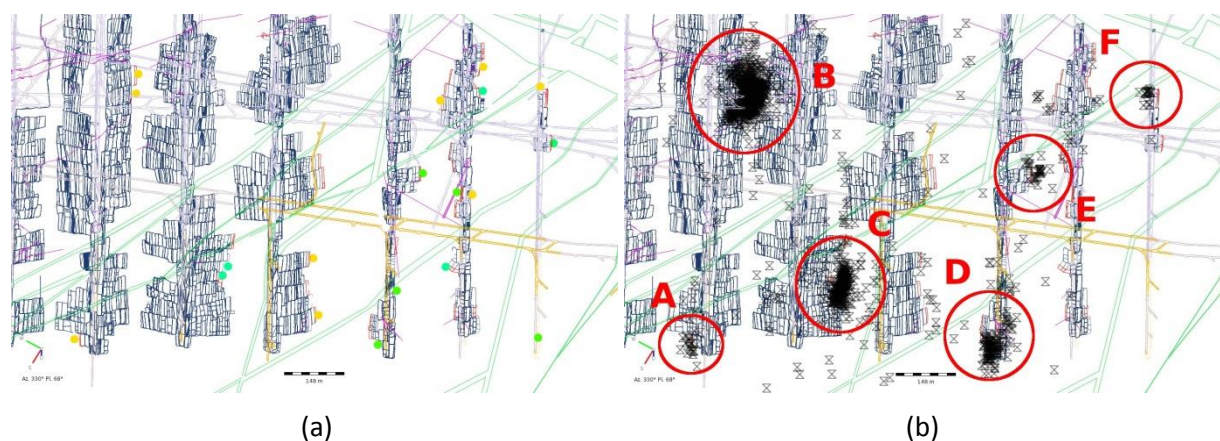
Figure 5   Plan view of a section of the mine with the production points (a); and all the events recorded during a 24 hour period (b)

The resulting clusters can be seen in Figure 6. The clusters are coloured to indicate whether they are associated with production or not. The light clusters are associated with production while the darker clusters are not associated with production. It is also noted that the observed clusters B-F are produced by the clustering algorithm. The clustering algorithm was also able to obtain the hidden temporal clusters. Even though the events in cluster A were spatially clustered, it was found that they were not temporally clustered and thus the clustering algorithm did not cluster them.

(a)   (b)
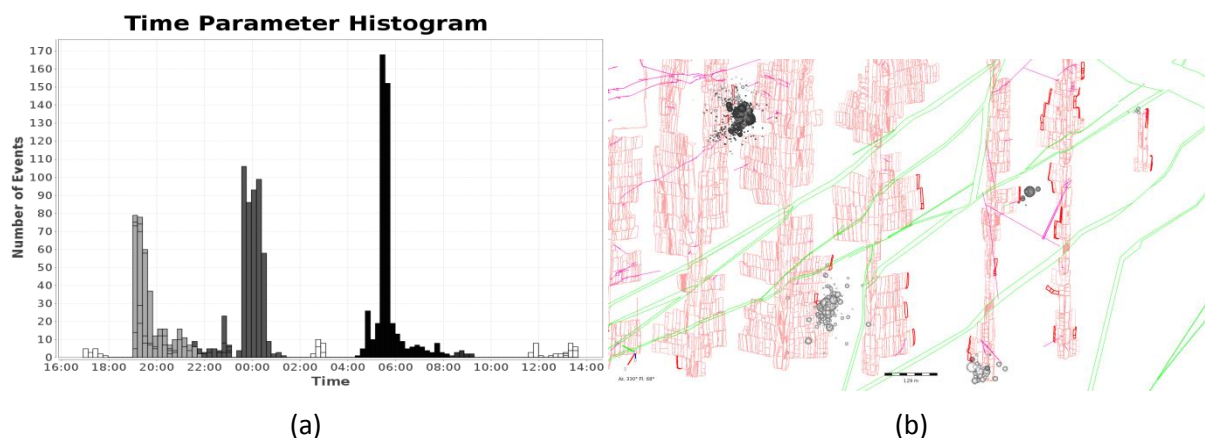
Figure 6   Histogram of the event times showing the clusters related to blasting (a); the clusters produced by the DCA as indicated by spheres sized by local magnitude (b)

This case demonstrates the ability of the clustering algorithm to identify clusters which are associated with production as well cluster that are not associated with production. The automatic clustering will augment the human analysis by highlighting the production versus non-production related clusters.

In another example, mine wide spatial and temporal clustering was performed and an anomalous region was found. This smaller region of the mine can be seen in Figure 7. The results of the clustering in this area show three prominent clusters indicated by shades of grey. The light grey cluster is observed around blasting time and is likely associated with production. The grey cluster indicates anomalous events recorded around midnight. A further black cluster was observed during the early morning. This cluster includes a m1.3 event which was recorded at 5:38 am. The event is marked in Figure 7 with a vertical line in the histogram. The same clustering analysis was performed on the preceding day to test whether the anomalous cluster, observed at midnight, may indicate an instability in the rock mass. The results of this clustering can be seen in Figure 8.

(a)   (b)
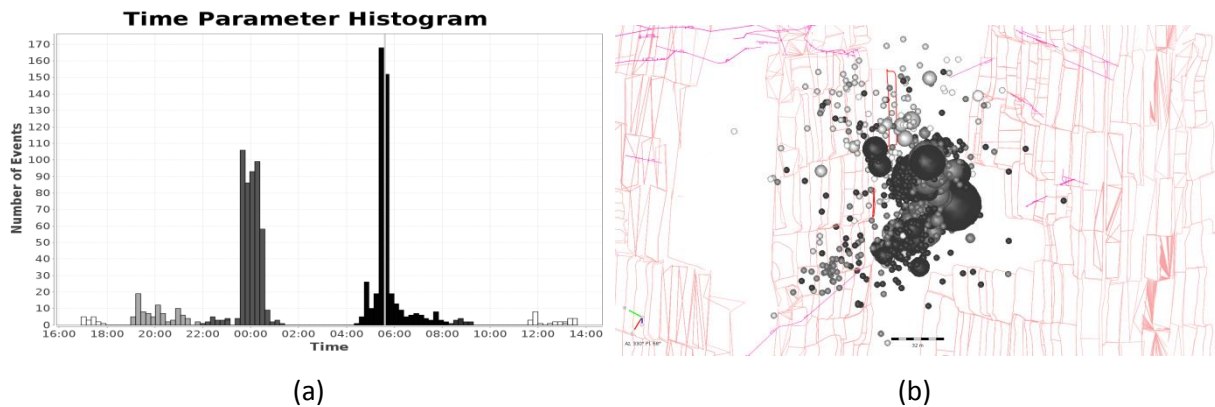
Figure 7 **Histogram of the event times showing the three prominent clusters indicated by shades of grey (a); the vertical line indicates the time of the large event; event plot showing the three prominent clusters (b)**

The strong blast time cluster is the most prominent cluster observed in Figure 8. A cluster of events is observed around midnight but it is not as prominent as the one seen in Figure 7. This suggests that previously unseen clusters observed outside of blasting time can indicate instabilities in the rock mass. This could be implemented as an automatic alarm to prompt human investigation into the anomalous cluster. Another advantage is that, historically, regions of monitoring are generally set up manually before-hand, while spatial filtering automatically emerges from the clustering.


(a)   (b)
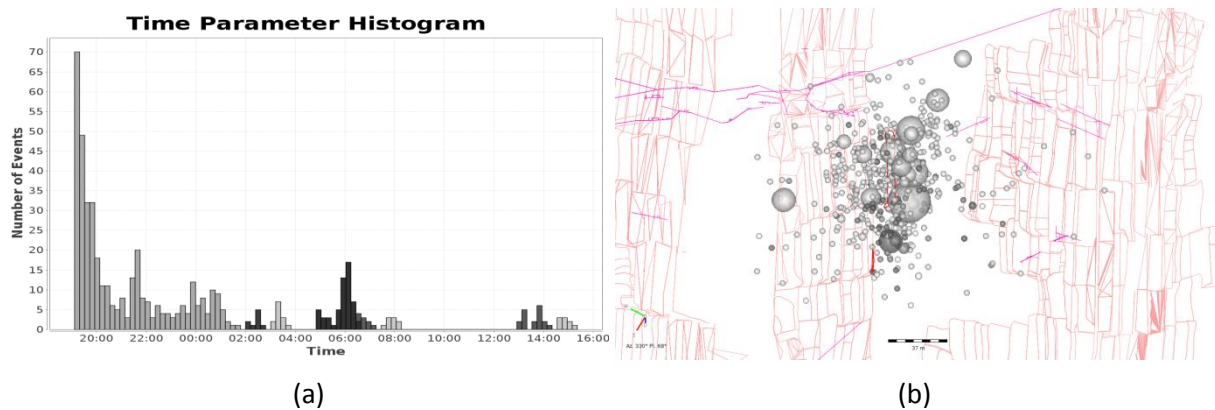
Figure 8 **Histogram of the event times showing the prominent blast time cluster indicated in light grey (a); a cluster of activity is observed around midnight, indicated in dark grey; event plot of the clusters (b)**

3.2   Spatial-E_S/E_P clustering

The effect of comparing clusters from different time scales was also investigated. This was done by finding a large event and clustering the events recorded during the preceding day as well as the events recorded during the preceding week. In the example shown in Figure 9, the clustering was performed on the spatial

parameters as well as the E_S/E_P ratio. For this mine, the predominant failure mechanism for large events is shear type events which typically have high E_S/E_P ratios. The cluster indicated by hourglasses is the result of clustering the week preceding the large event while the clusters indicated by spheres are the results of clustering the preceding day. An interesting observation is the presence of a high E_S/E_P ratio cluster observed during the preceding day but not during the preceding week. Under normal spatial clustering this may have been missed but here the multi-dimensional clustering highlights this new subcluster. This suggests that it may be possible to automatically identify instabilities by observing the changes in the clustering results on different time scales.

(a)   (b)
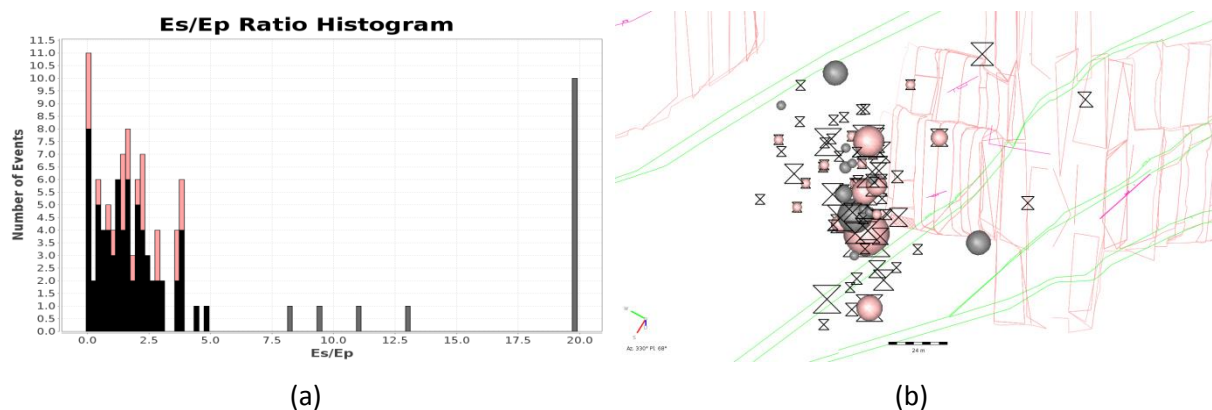
Figure 9   **Histogram of the E_S/E_P ratio of the events (a). The cluster produced by the long-term (week preceding the large event) clustering is indicated by hourglasses while the spheres correspond to the short-term (day preceding the large event) clusters. The event plot shows the results of the clustering (b)**

Preliminary work on how to interpret such clusters into a value of seismic hazard was investigated. At mines where large events usually have shear type mechanisms, it is probable that using the E_S/E_P ratio as a clustering parameter could indicate seismic hazard. To test this, two months of seismic data from a region of a mine was clustered using spatial, temporal and E_S/E_P ratio parameters. The results of the clustering can be seen in Figure 10. Events exceeding m1.0 are indicated by vertical markers. There were eleven large events which were grouped into eight distinct large event occurrences. This would mean a random probability of an event exceeding m1.0 on any given day of 14%. The clusters are grouped according to the mean E_S/E_P ratio as follows: black clusters indicate high ratios (>8), white indicates low ratios (<5) and grey indicates ratios between the high and low values. By looking at high E_S/E_P clusters with more than ten events, there is a 88% chance of a large event occurring within one day of finding these precursory clusters (success rate), a 13% chance of missing a large event and a 42% chance that there won't be a large event following the precursory cluster. The success rate of using this technique is 6.3 times better than randomly guessing each day whether there will be a large event. This suggests that clustering on E_S/E_P ratio in this region may add value to seismic hazard analysis.
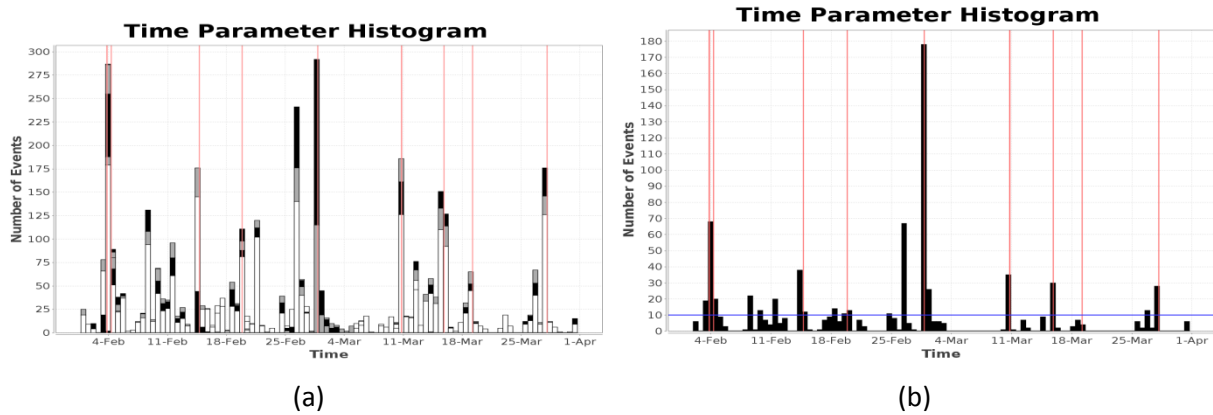
(a)   (b)

Figure 10 The results of clustering the spatial, temporal and E_S/E_P ratio parameters (a). The E_S/E_P ratio clusters are indicated by colours where black indicates high E_S/E_P events, white indicates low E_S/E_P events and grey indicates the events in between the high and low values. The vertical markers indicate large events (m>1.0). The high E_S/E_P clusters (b) are shown together with the event count threshold (horizontal line)

3.3 Spatial-energy index clustering

Prior to a large event in a specific region of the mine the classic temporal trend of falling EI and accelerating CAV was observed (Figure 11). This area was selected to test if spatial-and-source clustering would be useful. Here spatial-and-EI clustering for a period leading up to the time of high EI (9 September) showed a number of low EI clusters even though the general temporal trend was indicating a slow increase in EI. These low EI events were mostly around the eastern panels (Figure 12) but contained few events. Only the top three clusters are indicated on the spatial map, so as not to clutter the picture too much.

From then, leading up to the large event the picture changes quite dramatically with a single low EI cluster emerging on the eastern panels. Figure 13 shows the histogram of the clustering along with the spatial map which is dominated by a low EI cluster on the east. The clustering is in agreement with the time history trend but has the advantage that it can be identified automatically more easily and could offer a better indication of the spatial extent of the instability rather than a manually pre-defined region.
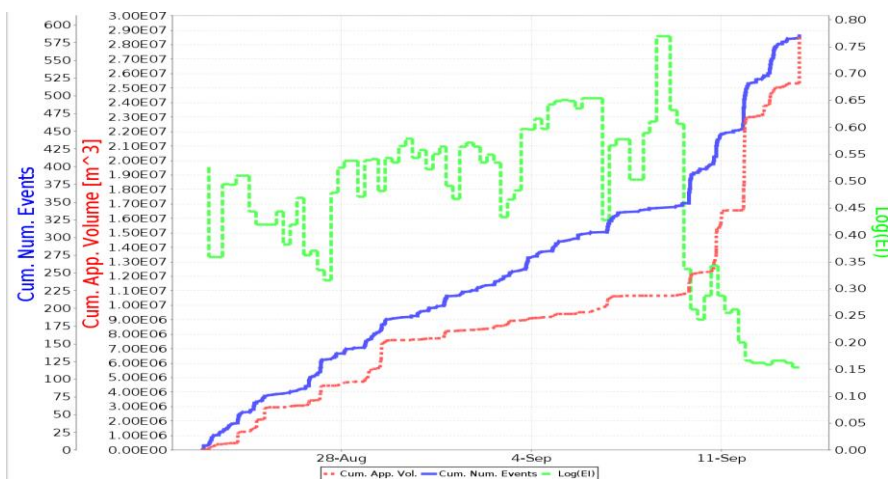

Figure 11 The time history of EI and CAV from an area around the stopes for a period leading up to a large event. EI is on a slow upward trend up until around the 9th of September but a strong drop with accelerated CAV occurs from then until the large event
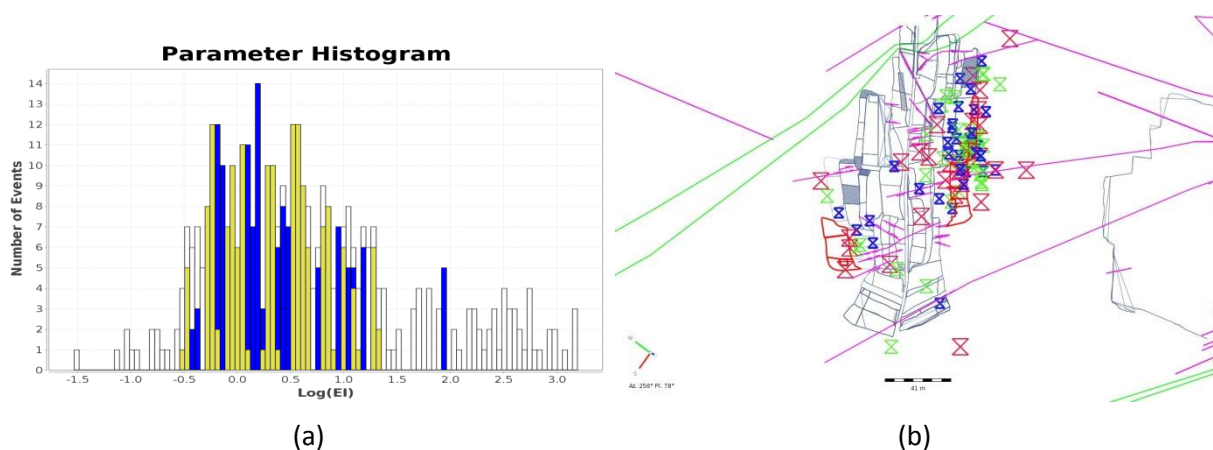
(a)

(b)

Figure 12 Histogram of the clustering results (a). The colouring has been chosen to differentiate between the clusters with the white cluster showing the unclustered events. The top three clusters are shown as hourglasses (b)
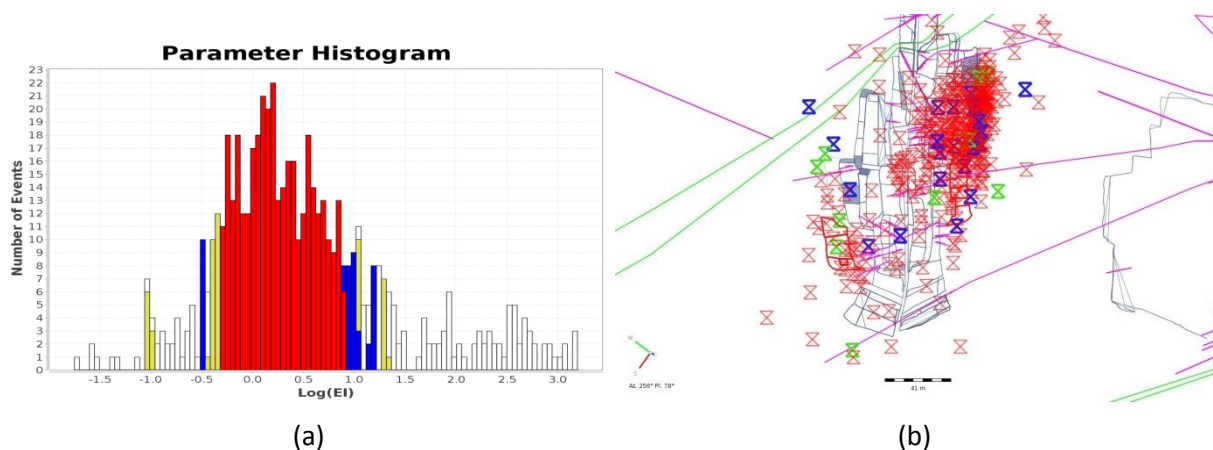

(a)

(b)

Figure 13 Histogram of the clustering results (a). The dominant low EI cluster indicated. The clusters are shown (b) with the dominant low EI cluster located at the eastern panels

4   Conclusion

Short-term hazard assessment has been around for a number of years in the deep South African gold mines. A number of improvements have been added over time. This paper describes investigations into the use of clustering methods as another tool which could be used for automatic detection of short-term instabilities.

Three different types of clustering were tested on a mine, namely spatial-and-temporal, spatial-and-E_S/E_P and spatial-and-EI clustering, using the DBSCAN method of clustering.

The spatial-and-temporal clustering was able to detect an anomalous cluster of seismicity a few hours after a production blast which culminated in a large seismic event. Since this cluster was late at night, it is unlikely it would have been detected by a human, unless the mine has dedicated staff watching the seismicity on a 24 hour, seven days a week basis. Automating this type of clustering would help in highlighting the instability ahead of time with the possibility of an early alarm. It has the added advantage of being able to identify these anomalous clusters throughout the mine without the user needing to setup spatial regions of interest.

On a mine where the predominant failure mechanism for large events is shear, the study investigated the benefits of spatial-E_S/E_P clustering. Testing this over a period of two months showed that clusters of high E_S/E_P events were observed prior to large events 88% of the time. Although a few large events were missed

and a few false alarms were present, the success rate was well above the random likelihood of a large event occurring.

The third clustering test involved looking for the more classical drop in EI precursory trend. Here spatial-EI clustering highlighted an area of low EI a few days before a large seismic event. Again the advantage is that predefined monitoring regions are not necessary and automatic alarming on these clusters is possible.

The three clustering types tested have shown that automatic detection of precursory trends prior to large events is possible. Using all three clustering types and automatically notifying the relevant mining personal could be new tools in the short-term seismic hazard monitoring arsenal.

Acknowledgement

References

Brink, A & Mountfort, P 1984, 'Feasibility studies on the prediction of rockbursts at Western Deep Levels', in NC Gay & EH Wainwright (eds), *Proceedings of 1st International Congress on Rockbursts and Seismicity in Mines*, South African Institute of Mining and Metallurgy, Johannesburg, pp. 317-325.

Ester, M, Kriegel, H, Sander, J & Xu, X 1996, 'A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise', in E Simoudis, J Han & U Fayyad (eds), *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Association for the Advancement of Artificial Intelligence, Portland, pp. 226-231.

Hills, P, Rebuli, D & Lynch, R 2013, 'Micro-Seismic Monitoring at the Tasmania mine, Beaconsfield, Tasmania', in A Malovichko & D Malovichko (eds), *Proceedings of 8th International Symposium on Rockbursts and Seismicity in Mines*, Geophysical Survey of Russian Academy of Sciences, Obninsk, Mining Institute of Ural Branch of Russian Academy of Sciences, Perm, pp. 485-494.

Mendecki, AJ 1997, *Seismic Monitoring in Mines*, Chapman & Hall, London.

Milligan, G & Cooper, M 1988, 'A study of standardization of variables in cluster analysis', *Journal of Classification*, vol. 5, pp. 181-204.

Rebuli, D & van Aswegen, G 2013, 'Short term seismic hazard assessment in S.A. gold mines', in A Malovichko & D Malovichko (eds), Proceedings of 8th International Symposium on Rockbursts and Seismicity in Mines, Geophysical Survey of Russian Academy of Sciences, Obninsk, Mining Institute of Ural Branch of Russian Academy of Sciences, Perm, pp. 323-331.

van Aswegen, G 2005, 'Routine seismic hazard assessment in some South African mines', in Y Potvin & M Hudyma (eds), *Proceedings of Sixth International Symposium on Rockburst and Seismicity in Mines*, Australian Centre of Geomechanics, Perth, pp. 437-444.