Unsupervised clustering of mining-induced microseismicity

Himanshu Barthwal[1], and Robert Shcherbakov[1,2]
[1]Department of Earth Sciences, University of Western Ontario, Canada.
[2]Department of Physics and Astronomy, University of Western Ontario, Canada.

Summary

We propose the application of unsupervised clustering to group microseismic events into different classes directly from the waveform data such that the events in a specific class have similar source mechanisms. Our method has three main steps, first using spectral decomposition to separate the source terms from the path-receiver contributions in the observed amplitude spectra of events occurring in spatially dense clusters. Second, reducing the number of features from the source spectra using independent component analysis. Third, applying the k-means algorithm to the reduced feature matrix to obtain event clusters. To test our method, we generate synthetic waveform data using the receiver network and the recorded microseismic event locations in an underground potash mine in Saskatchewan. We achieve reliable results for synthetic cases with well-defined moment tensor clusters. This method can offer an assessment of source types of large microseismic populations as often encountered in induced seismicity.

Method

We use the spectral decomposition technique (Shearer et al., 2006; Trugman and Shearer, 2017) that aims to separate the source spectra from the path and site terms. It is based on the premise that for a dense event cluster, each event is recorded at multiple receivers, each receiver records many events so that each approximate event-station path is sampled many times. Thus, the amplitude spectra of events belonging to a spatially dense cluster recorded at multiple receivers can be used to obtain their relative source spectra. A feature vector for an event is thus obtained by concatenating the relative source spectra at the three components. For our data, the feature vector has 234 components.

Since clustering performance is usually poor in high-dimensional space, we reduce the number of features by applying independent component analysis (ICA). This method is often used for blind source separation by decomposing a signal into independent non-Gaussian components. We use plots of reconstruction errors to identify the number of components to keep for further processing. We use 10 components obtained from ICA to obtain a reduced feature matrix. Next, we apply the k-means clustering algorithm to this reduced feature matrix. We use a plot of the mean Silhouette coefficient for different numbers of clusters to determine the optimal number of clusters in the data. It is a measure of the proximity of a point in one cluster to points in the neighboring clusters. It ranges from -1 to +1 with +1 showing that the point is far from the points in other clusters, 0 indicating that it is near or on the decision boundary between two neighboring clusters, and a negative value showing the wrong cluster assignment.

Synthetic data

We base our synthetic example on the acquisition setup of a microseismic monitoring system in an underground potash mine located near Saskatoon, Saskatchewan. The mining activity induces microseismicity which is recorded by 12 in-mine broadband seismometers with depths varying from 0.99 to 1.05 km below ground level (Figure 1). The sampling rate is 200 Hz. The data recorded between 1 March and 30 June 2021 have been previously used by Barthwal et al., 2024, to perform event locations and moment tensor inversion. Figure 1 shows the locations of all 1224 events (gray-filled circles). To generate synthetic data, we compute Green's functions using the Computer Programs in Seismology package (Herrmann, 2013) for the 172 event locations (shown as black-filled circles in Figure 1) with a 1D velocity model representing mine geology (taken from Barthwal et al., 2024). These Green's functions are combined with 1000 moment tensors to compute synthetic seismograms for 1000 events.

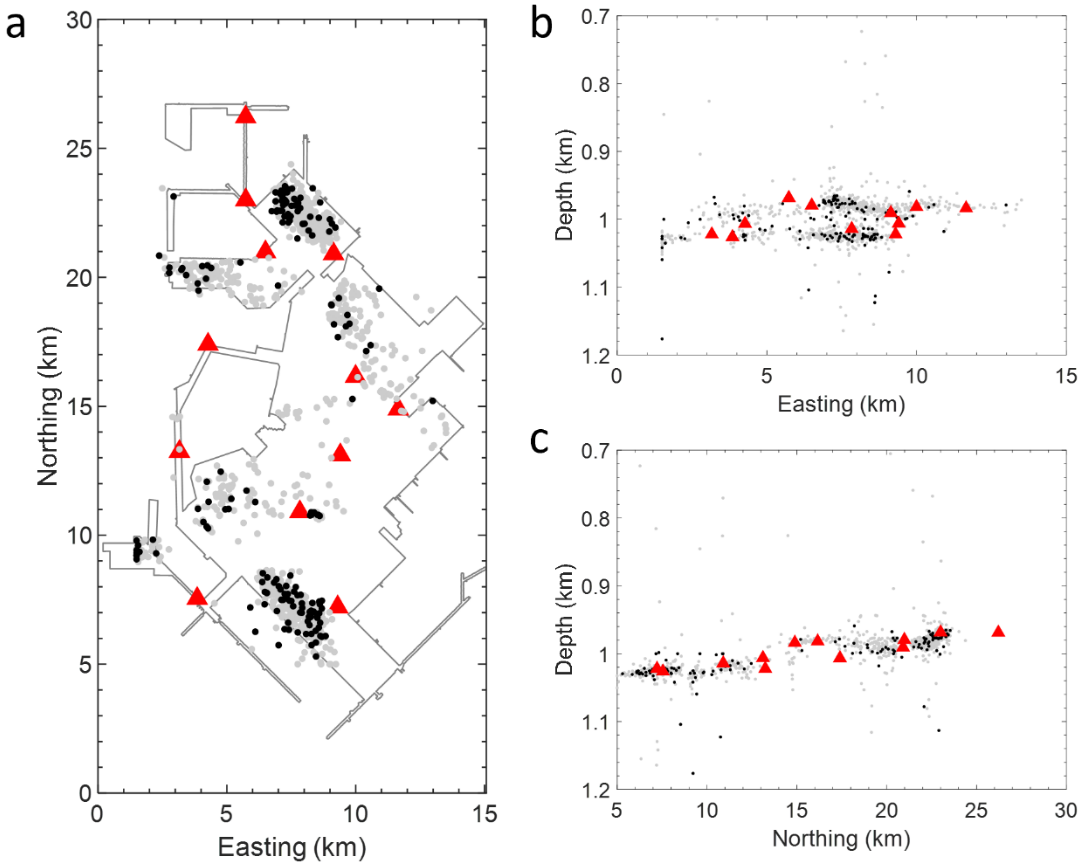

Figure 1. Event locations and underground receivers are shown by filled gray circles and red-filled triangles, respectively. The events shown with black color are used for synthetic tests. (a) Map view. Mine layout is shown in gray color. (b) East-West cross-section. (c) North-South cross-section.

Results

Figure 2 shows the strike, dip, and rake (Figure 2a), and the non-double-couple (non-DC) components in the Hudson plot (Figure 2b) that are used to generate 1000 instances of moment tensors. Figure 2c shows 172 event locations color-coded with the first 172 source mechanisms according to the three colors for the clusters shown in Figure 2b. The source mechanisms have been distributed randomly with no location dependence. This is done on purpose to test the ability of the proposed method to recover underlying groups based on source mechanisms irrespective of their locations. The double-couple (DC), isotropic (ISO), and compensated linear vector dipole (CLVD) groups observed in the Hudson plot have 337, 332, and 331 events, respectively. Figure 2d shows the mean Silhouette coefficient computed for different numbers of clusters. A maximum is seen for 3 classes. The clustering result using 3 clusters is shown in Figure 2e. The three groups are well resolved in the Hudson plot. Since clustering does not provide any labels, the resulting clusters are labeled (using different colors) based on the maximum number of events that are common to the synthetic groups. We get a high accuracy greater than 94% for all three synthetic classes.

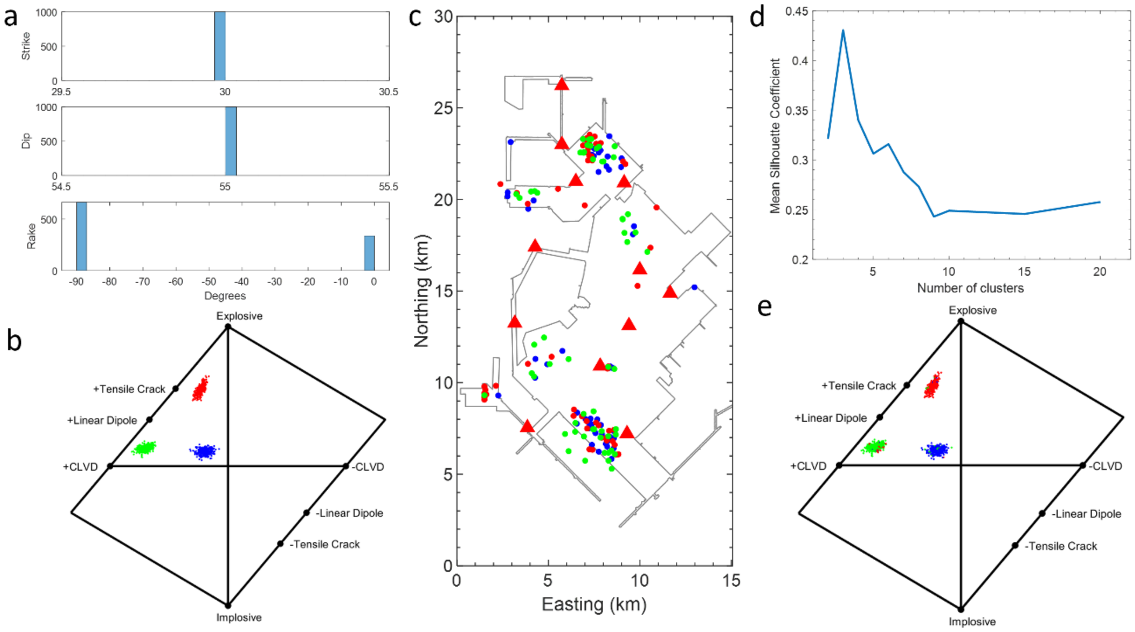

Figure 2. (a) Fault plane orientations (strike, dip, rake) to generate moment tensors. (b) Hudson plot showing the three groups with dominant DC (blue), ISO (red), and CLVD (green) components. (c) Map view of 172 events for which Greens' functions are computed. Colors represent the assigned moment tensors corresponding to the three clusters shown in Figure 2b. (d) Silhouette score computed for different numbers of clusters specified in the k-means algorithm. (e) Clustering results using the proposed method. The colors have been assigned based on the maximum common events between the synthetic groups and the clustering results.

Conclusions

We tested the feasibility of an unsupervised clustering approach to group events into different classes using their relative source spectra. Synthetic examples show that the method works well if there is some structure present in the data and can group events with similar source mechanisms together irrespective of their locations. This method is completely data-driven and does not make any assumptions about the physics of the source. Thus, it can be particularly useful for clustering the induced seismicity which often shows diverse source mechanisms including significant non-double-couple components.

Acknowledgments

We thank Nutrien Ltd. for the permission to use the data and show the results. This research has been supported by NSERC Discovery Grant.

References

Barthwal, H., M. van den Berghe, and R. Shcherbakov (2024), Microseismic event locations and source mechanisms using dominant guided waves recorded in an underground potash mine, GEOPHYSICS, 89, B51-B63. https://doi.org/10.1190/geo2023-0359.1

R. B. Herrmann (2013), Computer programs in seismology: An evolving tool for instruction and research, Seismological Research Letters, 84 (6), 1081–1088.

Shearer, P. M., G. A. Prieto, and E. Hauksson (2006), Comprehensive analysis of earthquake source spectra in southern California, J. Geophys. Res., 111, B06303, doi:10.1029/2005JB003979.

Trugman, D. T., and P. M. Shearer (2017), Application of an improved spectral decomposition method to examine earthquake source scaling in Southern California, J. Geophys. Res. Solid Earth, 122, 2890–2910, doi:10.1002/2017JB013971.