Quantifying Fracture Networks Inferred From Microseismic Point Clouds by a Gaussian Mixture Model With Physical Constraints

S. H. McKean1, J. A. Priest1, J. Dettmer2, and D. W. Eaton2

1Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada, 2Department of Geoscience, University of Calgary, Calgary, Alberta, Canada

Abstract Microseismicity is generated by slip on fractures and faults and can be used to infer natural or anthropogenic deformation processes in the subsurface. Yet identifying patterns and fractures from microseismic point clouds is a major challenge that typically relies on the skill and judgment of practitioners. Clustering has previously been applied to tackle this problem, but with limited success. Here, we introduce a probabilistic clustering method to identify fracture networks, based on a Gaussian mixture model algorithm with physical constraints. This method is applied to a rich microseismic data set recorded during the hydraulic fracturing of eight horizontal wells in western Canada. We show that the method is effective for distinguishing hydraulic-fracture-created events from induced seismicity. These fractures follow a log-normal distribution and reflect the physical mechanisms of the hydraulic fracturing process. We conclude that this method has wide applicability for interpreting natural and anthropogenic processes in the subsurface.

Plain Language Summary This paper presents a method for inferring fractures from numerous small-magnitude earthquakes (microseismicity) recorded during a hydraulic fracturing program in an oil and gas reservoir. This method probabilistically clusters microseismic events into a network of fractures after reclassifying them into stages and applying physical constraints. It is superior to previously published clustering algorithms for this purpose. It also allows us to distinguish fractures created by hydraulic fracturing from those triggered on preexisting faults (induced seismicity). This information is essential for assessing both natural and triggered earthquakes and helps us understand deformation processes in the Earth.

1. Introduction

Fracture networks play a fundamental role for subsurface fluid transport in oil and gas reservoirs, volcanoes, fault systems, and aquifers. Many types of discrete-fracture-network models have been developed to understand these processes. Some examples include models of contaminant transport (Berkowitz, 2002), hydro-shearing for geothermal energy (Tezuka & Niitsuma, 2000), blockcaving for mining (Hudyma, 2008), and hydraulic fracturing (HF) for unconventional oil and gas (Maxwell, Mack, et al., 2015). However, the general inaccessibility of subsurface-fracture networks means that inference methods for characterization and analysis are required.

Slip activation of fractures and faults can generate seismic waves that can be used to characterize natural and anthropogenic fluid transport in the subsurface. This microseismicity can provide important clues about the location, density, and orientation of fractures (Eaton, 2018). Microseismic point clouds typically have noisy, complex, and incomplete spatiotemporal expressions. This complicates their use in investigating subsurface fracture networks. Most attempts to infer fracture characteristics rely on heuristic pattern recognition, but the large number of events in modern microseismic catalogs quickly overwhelms these approaches. This problem is compounded by the multivariate attributes of microseismic data, which tends to confound simple analytical methods such as cross plots.

This study investigates HF-created microseismicity and induced seismicity (IS; Shapiro et al., 2006; Eaton, 2018). Hydraulic fracturing creates new fractures and dilates natural fractures by injecting fluids at high pressure through multiple discrete stages in a wellbore (Hubbert & Willis, 1972; Smith & Montgomery, 2015). The creation of new fractures is initiated by a negative effective stress that exceeds the tensile strength of
intact rock (Johann et al., 2016). The propagation of this fracture then proceeds toward the maximum horizontal stress in accordance with fracture mechanics. The propagation is governed by the net pressure, elastic properties of the surrounding rock and the fracture toughness of the rock in front of the propagating fracture (Valko & Economides, 1995). The failure of natural fractures is also an effective stress-driven process, but it can occur in either shear or tensile fracturing mode when the frictional resistance and effective stresses surrounding a natural fracture are overcome (Rutledge & Phillips, 2003). Microseismicity is a key tool for understanding the physics of HF through the failure of intact rock and natural fractures (Warpinski, 2013).

Algorithmic approaches that employ unsupervised machine learning are ideally suited to multivariate inference in complex, high-dimensional data sets without labels. These approaches identify patterns and groups directly from similarities in attributes (The Royal Society, 2017), creating clusters. Numerous clustering methods have been applied to microseismic data. Some examples include waveform analysis (Moriya et al., 2005; Skoumal et al., 2015; Tezuka & Niitsuma, 2000), matched-filtering (Caffagni et al., 2016; Eaton & Caffagni, 2015), principal component analysis (PCA) with graph theory (Vasudevan et al., 2010), the k-means algorithm (Eaton et al., 2014), and density-based clustering (Cesca et al., 2013). Unsupervised machine learning techniques have also been applied for automated microseismic data arrival picking (Chen, 2018a) and fast waveform detection in microseismic imaging problems (Chen, 2018b).

In this paper, we introduce a new clustering method that employs a Gaussian mixture model (GMM) with physical constraints. These include a spatiotemporal-based separation of events into stages followed by the application of a diffusion-based model (Shapiro, 2015) to separate HF-created microseismicity from IS. The method is applied to a rich microseismic data set recorded during a hydraulic-fracturing treatment. We show that the application of hybrid analytical-clustering methods holds significant promise for characterizing fractures in the subsurface and investigating IS. This is one of the first applications of such a method for fracture quantification and we show its superior performance relative to application of GMMs only.

2. Data and Methods

The data set in this study (Figure 1) consists of 17,540 microseismic events observed during a HF treatment of eight horizontal wells (Figures 1a and 1b). The wells are located at a depth of ~2,500 m within the Duvernay formation, a prolific unconventional oil and gas formation in Western Canada. The maximum horizontal stress roughly trends east-west in the study area, which is in a strike-slip tectonic regime (Anderson, 1951). The HF program stimulated stages in multiple wells simultaneously using an industrial practice called zipper fracking (Eaton, 2018). This causes concurrent microseismic events in multiple wells and further challenges event interpretation. The HF program consisted of 153 stages completed during two time periods (Figure 1c). Wells 1 to 4 were completed on days 0 through 17 and wells 5 to 8 were completed on days 26 through 42. The microseismic signatures are illustrated by the overlapping colors (representing different wells) in Figure 1c. Between 430 and 3,000 m³ of fluid and 45 to 300 tons of proppant were injected into each stage. Several induced seismic events are visible east of the wells (Figure 1a). The observed microseismic events are characterized by reported moment magnitudes (MW) in the range ~2.4 to 2.3. The magnitudes were estimated based on low-frequency spectral amplitudes, and it is likely that events exceeding MW 1 are above the saturation limits of the geophones (cf. Eaton et al., 2018). This limitation has no effect on the clustering analysis considered here.

The microseismic data set is defined by spatial coordinates (R), origin time (T), Mw, moment tensor (MT) solution. Each event in the catalog is also assigned to a well and stage. To process the data, events below the magnitude of completeness (Mc) are removed. The Mc is the lowest magnitude that a specific geophone array records completely. It is determined using a combination of the binned maximum-likelihood (Aki, 1965; Bender, 1983; Freedman & Diaconis, 1981), maximum curvature, and goodness-of-fit methods (Wiemer & Wyss, 2000) with a cross-validated approach (Kuhn et al., 2008). Removing events below the Mc provides an unbiased data set with 11,978 events. Additional details on the data processing are provided in Text S1 in the supporting information. The data set has event location uncertainty that ranges from 0.1 m for large-magnitude events to 900 m for low-magnitude events near the study area boundary. This, along with the uncertain HF initiation point, creates variability in the clustering method results. A bootstrap resampling (Text S2, Efron (1982)) of the location uncertainty and HF initiation point uncertainty (Williamson, 1999) quantifies the mean (M) and standard deviation (s.d.) of 10,000 realizations for key metrics in this study.
GMM algorithms rely on the multidimensional distance (or dissimilarity) between data attributes. Consequently, the attributes used in the method strongly affect its results. We evaluate the importance of each attribute and select an appropriate subset of attributes using correlation and PCAs. First, a

Figure 1. Plan view (a) and side view (b) of the microseismic data set, colored by operator well classification and sized by magnitude. Well trajectories (black lines), HF stages (red points), and well numbers (black text in white boxes) are shown. The microseismic events are shown by circles, with colors representing the well, and the circle size representing magnitude. A magnitude-time plot (c) shows the sequencing of the two portions of the HF completions. Vertical lines are colored by well and represent the duration of each individual stage. The duration of each stage is also illustrated as a heat map (d).
nonparametric correlation (Spearman, 1904) is conducted, as opposed to a Pearson correlation (Pearson, 1897), because the data are nonnormal (Rahman & Govindaraju, 1997; Razali & Wah, 2011; Shapiro & Wilk, 1965; Wei & Simko, 2013). Then, PCA is used to visualize the influence of each attribute on the data variance and group-similar attributes together in reduced-dimensionality space (Hotelling, 1933; Pearson, 1901). The application of correlation and PCAs shows that the $R$ and $T$ attributes are more influential than the MT and $M_{60}$ attributes. Based on this, a lower-dimension data set with $R$ and $T$ attributes only is considered in the clustering method. Creating this subset is shown to improve the performance of the clustering algorithm. Thus, the use of correlation and PCA (Text S3) is an essential step in the implementation of the algorithm. Additional details on correlation and PCAs and their results are provided in Text S3. Beyond attribute selection, the PCA results also suggest that fracture orientation and spatiotemporal proximity of the microseismic events to the fracturing stage play a role in the generated magnitude.

We compare and validate the GMM method against three widely used deterministic clustering algorithms: K-medoids (MacQueen, 1967; Kaufman & Rousseeuw, 2009; Hennig, 2013), agglomerative clustering (R Core Team, 2014; Ward, 1963), and density-based spatial clustering with noise (Sander et al., 1998; Hahsler & Piekenbrock, 2017). Validation statistics based on the within-cluster and between-cluster distances are applied to evaluate this comparison (Califskå & Harabasz, 1974; Charrad et al., 2014; Kaufman & Rousseeuw, 1987). Details on clustering are discussed in Text S4. The comparison shows that the GMM method reveals lineations in regions of high point cloud density, where deterministic algorithms struggle with noisy point cloud data.

The GMM algorithm is a probabilistic clustering method (Banfield & Raftery, 1993; Celeux and Govaert, 1995, Fraley & Raftery, 2006) that assigns multivariate Gaussian distributions via an expectation-maximization algorithm (Dempster et al., 1977) with the Bayesian information criterion (Schwarz, 1978). The GMM algorithm can address overlapping clusters, uncertainty in point locations, and high densities of microseismic observations to a degree. The use of the Bayesian information criterion results in a parsimonious solution that approximates Bayesian evidence, optimizing the trade-off between data fit and complexity (i.e., the number of clusters). The GMM algorithm has been successfully applied to studies in hydrology (Kim et al., 2014), energy (Kazor & Hering, 2015), and geochemistry (Ellefsen et al., 2014). It also has the advantage of defining a confidence ellipse for each cluster. The geometric properties of each ellipse are directly interpretable, avoiding the need to apply additional techniques such as the convex hull method (Barber et al., 1996). An ellipsoid is an ideal representation of a hydraulic fracture, which tends to have dimensions of the order of $10^2$ m (length), $10^4$ to $10^5$ m (height), and $10^{-2}$ m (aperture; Nelson, 2001). The ratio of the three principal axes in each cluster quantifies its planarity and thus the confidence that each cluster is a fracture. Near-spherical clusters that contain a small number of microseismic events are classified as spurious and removed from further interpretation. In addition, a series of simple physical constraints are introduced to segregate the point cloud into groups and to improve the GMM method’s ability to identify fractures.

3. Results and Discussion

The GMM method is first applied to $R$ and $T$ attributes from the preprocessed data set, resulting in the identification of 25 clusters (Figure 2). This unconstrained application reveals a cluster of induced events east of the wells (identified by the gray clustered points in Figure 1) but is unable to meaningfully cluster hydraulic fractures. Most clusters trend parallel to the horizontal portion of each well and close to the injection stages, where the density of microseismic events is highest. But, they generally fail to capture the geometry of hydraulic fractures, which are expected to be elongated clusters oriented perpendicular to the minimum-stress axis. A more granular analysis with stage-level separation is therefore required for a meaningful investigation of microseismicity.

The stage-level separation improves clustering performance by dividing the data set into 152 stages and applying a new classification algorithm to associate each microseismic event to a completion stage, resulting in up to 310 events ($M$: 310, s.d.: 2). This algorithm applies a Euclidian distance in space and time (see Text S5). Note that classification prior to this part of the analysis was based on operator classification which considers only the timing of microseismic events relative to the start of a completion stage. However, HF is a fluid-driven diffusive process and it has been demonstrated that microseismic events can continue for hours and days after the end of a particular stage (Johann et al., 2016; Shapiro, 2015; Shapiro et al., 2006).
Separating stages by origin time alone can result in erroneously classified events that are simply a continuation of the HF process after an injection stage ends. Use of both time and distance reclassifies ~25% of the data set relative to the time-only classification. This significantly improves the clustering performance, as illustrated by a portion of the data set from well 3, stage 2 (w3s2) that is provided as Data Set S1. The original data for this single-stage portion (Figure 3a) contain 310 events and reveal seven clusters, whereas the reclassified data (Figure 3b) contain 216 events ($M$: 203, s.d.: 21) and reveal four clusters. The reclassification of the distant event in the northeast tightens the noisy cluster (red in Figures 3a and 3b) and eliminates a previously identified lineation (dark blue, Figure 3a) to provide a more parsimonious solution.

Figure 2. Orthographic projection of the clustering results from the application of the Gaussian mixture model algorithm to the preprocessed microseismic data set. The clusters are differentiated by color, with (a) plan view showing two clusters of induced seismicity east of the wells in light and dark green, (b) northing side view with most clusters at higher elevation than the wells, and (c) easting side view. The black arrows show the approximate orientation of the maximum principal stress in the study area.
Figure 3. Plan view of the Gaussian mixture model method applied to events from w3s2 for (a) the original data set ($n = 310$), and (b) the reclassified data set ($n = 216$). Reclassification of the distant event in the northeast modifies the red cluster (3_2_HF1) and eliminates 3_2_HF5 (a) in favor of a larger, noisy cluster. The wells (black lines) and hydraulic fracturing stages (red dots) are also shown. HF = hydraulic fracturing.

Figure 4. Plan view of the clustering results from the w3s2 data set after outlier detection. Part (a) shows the five clusters identified in the hydraulic fracturing (HF) zone of influence. Part (b) shows three clusters identified as induced microseismicity.
The events associated with each stage still contain both HF-created and induced events. Therefore, a physical constraint reflecting HF physics is required to separate these two types of events. Here, we apply an analytical diffusion-based model (Shapiro et al., 2006) to define a sphere of influence for each hydraulic fracture. This model (cf. Shapiro, 2015, pp. 167, eq. 4.3 & 4.4) relies on the approximates of the Perkins, Kern, and Nordgren semianalytical model of hydraulic fracture growth in an isotropic rock mass (Nordgren, 1972). The length of a constant-height \( h_f \) hydraulic fracture \( L \) is estimated based on the average injection rate \( q_i \) and fluid leakoff \( c_L \) into the formation, as detailed in equations (1) and (2) and Text S5. We assume a fracture height equal to the average thickness of the Duvernay formation (75 m) in the study area and a typical leakoff coefficient for shale (0.001 m/min; Munson, 2015).

Figure 5. Equal-area stereographic projections and density plot of poles showing the hydraulic fracturing-created clusters (a) and induced seismicity clusters (b) identified by the Gaussian mixture model algorithm after reclassification and application of physical constraints. Each contour line represents 0.5 standard deviations and the contour edge represents a 95% confidence interval. Normalized semi-log histograms illustrate the height (blue) and length (red) of the hydraulic fracturing-created clusters (c) and IS clusters (d). The empirical cumulative distribution functions (black lines) are compared with log-normal distributions (blue for height and red for length) that were fit using the maximum likelihood method.
The model predicts a mean fracture length of 379 m for w3s2 where fluid was injected at a rate of 14.9 m³/min for 390 min. This length is increased by 1 standard deviation to project a spherical region of influence of 712 m. This sphere physically represents an estimate of the stimulated reservoir volume (Maxwell et al., 2009), and it is used to separate HF-created events and induced microseismicity. The variance is a tuning parameter for the method, but 1 standard deviation seems to reflect the likely extent of a hydraulic fracture. Two standard deviations results in spheres of influence nearly 10 times the expected size of a hydraulic fracture. The application of the physical constraint separates the stage-segregated events into 148 HF-created groups (M: 146, s.d.: 0.5) and 38 IS groups (M: 37, s.d.: 1.2). This successful application of a diffusion-based spatiotemporal model to separate HF-created microseismicity and IS reinforces the idea that HF-created microseismicity is fluid driven and reflects the geomechanical response of unconventional reservoirs.

We then apply the GMM to each of the 186 groups of HF-created and IS groups resulting in the identification of 488 clusters (M: 484, s.d.: 9.6). Figure 4 presents the results for w3s2, with the results from the entire data set presented as Figures S13 and S14. In many cases, the HF-created events are clustered into single ellipsoids as a result of optimizing the number of clusters with the Bayesian information criterion. In other cases, multiple clusters (up to 11, M: 8, s.d.: 0.8) are identified. The orientation and dimensions of each cluster are quantified by a multidimensional 95% confidence ellipsoid. The covariance matrix of each ellipse represents the principal axes of the cluster, which quantify fracture length and height (Text S5).

Using the results, we now study the physical meaning of clusters. In an isotropic and homogeneous rock mass, hydraulic fractures propagate parallel to the maximum principal stress (Jaeger, Cook, and Zimmerman, 2007). Induced seismicity, in contrast, is expected to be localized within regions that are oblique to the maximum principal stress, where the shear stress is maximized (Ellsworth, 2013). Equal-area stereographic projections of the clustering results (Figure 5) show a significant difference of strikes of the HF-created (Figure 5a) and induced (Figure 5b) clusters. The HF-created fractures strike between 68° and 113° (bootstrap interquartile range), with a median of 90° (M: 89.6°, s.d.: 0.80°), parallel to the direction of the maximum principal stress. The IS clusters strike between 36° and 138° (bootstrap interquartile range), with a median of 63° (M: 61.1°, s.d.: 2.64°). Steep dips are observed for the majority of the clusters in this study, which is consistent with the slip-slip faulting regime (Anderson, 1951) that currently exists in the Duvernay (Schultz et al., 2017). Fractures with dips above 60° are lithologically bounded with a median height of 100 m (Figure 5c, M: 96.4 m, s.d.: 5.9 m). The fracture lengths (Figure 5d) range between 2 and 10 times the fracture height (M: 309 m, s.d.: 15.9 m). This is consistent with previous numerical simulations of fracture growth (Maxwell, Zhang, & Damjanac, 2015) and the assumptions of the Perkins, Kern, and Nordgren model approximation. The distribution of fracture heights and lengths are both well fit by log-normal distributions in this study (Figures 5c and 5d). This provides evidence that the results reflect a fracture network, since it has been established that fracture networks are self-similar with attributes generally following log-normal distributions (Eaton et al., 2014; Gutierrez & Youn, 2015; Nelson, 2001).

4. Conclusions

This study presents a new method for fracture network identification from microseismic point cloud data. Our method uses a GMM algorithm with physical constraints to probabilistically cluster events into hydraulically-created fractures and IS. The method is unbiased, automated, and provides a numerical approach for fitting elliptical planes in multidimensional spaces. It is computationally efficient and can be applied in real time. A HF program in an unconventional oil and gas reservoir is used as a case study. We demonstrate that our method is more effective than other deterministic clustering algorithms, revealing hundreds of fractures in a preprocessed microseismic data set with 11,978 events. The method successfully differentiates HF-created events from IS (fault slip) using a diffusion-based model. It is capable of extracting important information from microseismic data, which are inherently ambiguous due to the inaccessibility of the fractured region. The orientation and geometry of the identified fractures reflect the physical mechanisms of the HF process. This method, while not a panacea for clustering point cloud data, provides insights with
widespread implications for researchers and industry practitioners. For example, to identify weak fault planes during monitoring of mining or geothermal microseismicity. Comparing dense lineations to operationally expected microseismicity provides a new tool for fault identification and IS mitigation.

Acknowledgments
This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund and the Microseismic Industry Consortium. Our gratitude goes out to Anton Biryukov, John Evans, Nadine Igonin, and Volodymyr Vragov, and especially Hoda Rashedi for previous work on principal component analysis of microseismic data. The data and anonymized data subset are provided by two anonymous oil and gas producers. The data subset code (in R) can be found at this site (github.com/ScottHMcKean/mclustering).

References
Aki, K. (1965). Maximum likelihood estimate of b in the formula log N = a – bM and its confidence limits. Bulletin of the Earthquake Research Institute-University of Tokyo, 43, 237–239.
Anderson, E. M. (1951). The dynamics of faulting and dyke formation with applications to Britain. Hafner Pub. Co.
Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. Biometrics, 49(3), 803–821. https://doi.org/10.2307/2532201
Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS), 22(4), 469–483. https://doi.org/10.1145/235815.235821
Bender, B. (1983). Maximum likelihood estimation of b values for magnitude grouped data. Bulletin of the Seismological Society of America, 73(3), 831–851.
Berkowitz, B. (2002). Characterizing flow and transport in fractured geological media: A review. Advances in Water Resources, 25(8–12), 861–884. https://doi.org/10.1016/S0309-1708(02)00042-8
Caffagni, E., Eaton, D. W., Jones, J. P., & van der Baan, M. (2016). Detection and analysis of microseismic events using a Matched Filtering Algorithm (MFA). Geophysical Journal International, 206(1), 644–658.
Carter, T., & Harabasz, J. (1974). A complex method for cluster analysis. Communications in Statistics-Theory and Methods, 3(1), 1–27. https://doi.org/10.1080/03610927408827101
Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models Pattern recognition, 28(5), 781–793. https://doi.org/10.1016/0031-3201(94)00125-6
Cesca, S., Sen, A. T., & Dahm, T. (2013). Seismicity monitoring by cluster analysis of moment tensors. Geophysical Journal International, 190(3), 1813–1826.
Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. Journal of Statistical Software, 61(6). https://doi.org/10.18637/jss.v061.i06
Chen, Y. (2018a). Automatic microseismic event picking via unsupervised machine learning. Geophysical Journal International, 212(1), 88–102. https://doi.org/10.1093/gji/ggy420
Chen, Y. (2018b). Fast waveform detection for microseismic imaging using unsupervised machine learning. Geophysical Journal International, 215(2), 1185–1199. https://doi.org/10.1093/gji/ggy448
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B Methodological, 39(1), 1–22. https://doi.org/10.1111/j.2517-6617.1977.tb01600.x
Eaton, D. W. (2018). Passive seismic monitoring of induced seismicity: Fundamental principles and applications to energy technologies. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781131653554
Eaton, D. W., & Caffagni, E. (2015). Enhanced downhole microseismic processing using matched filtering analysis (MFA). First Break, 33(7), 49–55.
Eaton, D. W., Rafiq, A., Pedersen, P., & van der Baan, M. (2014). Microseismic expression of natural fracture activation in a tight sand reservoir. DFNE 2014, 265.
Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., & Rodriguez, G. (2018). Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors. Seismological Research Letters, 89, 1641–1651. https://doi.org/10.1785/0220180055
Ellson, B. (1982). The jackknife, the bootstrap, and other resampling plans, (Vol. 38). Siam: Society for Industrial and Applied Mathematics.
Ellenberg, J. J., Smith, D. B., & Horton, J. D. (2014). A modified procedure for mixture-model clustering of regional geochemical data. Applied Geochemistry, 51, 315–326. https://doi.org/10.1016/j.apgeochem.2014.10.011
Ellsworth, W. L. (2013). Injection induced earthquakes. Science, 341(6142). https://doi.org/10.1126/science.1225942
Fraley, C., & Raftery, A. E. (2006). MCLUST version 3: An R package for normal mixture modeling and model-based clustering. Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator: I: 2 theory. Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete, 57(4), 453–476. https://doi.org/10.1007/BF01025866
Gutierrez, M., & Youn, D. J. (2015). Effects of fracture distribution and length scale on the equivalent continuum elastic compliance of fractured rock masses. Journal of Rock Mechanics and Geotechnical Engineering, 7(6), 626–637. https://doi.org/10.1016/j.jrmge.2015.07.006
Hahsler, M., & Piekenbrock, M. (2017). dbscan: Density based clustering of applications with noise (DBSCAN) and related algorithms. R Package Version, 0–0.
Hennig, C. (2013). Fpc: Flexible procedures for clustering. R package version 2.1-5.
Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441. https://doi.org/10.1037/h0071325
Hubbert, M. King, and David G. Willis. (1972). Mechanics of hydraulic fracturing. Mem. - Am. Assoc. Pet. Geol. (18). 239-257.
Hudyma, M. R. (2008). Analysis and interpretation of clusters of seismic events in mines. University of Western Australia.
Jaeger, J. C., Cook, N. G. W., & Zimmerman, R. (2007). Fundamentals of rock mechanics (4th ed.). Maiden, MA: Blackwell Publishing. ISBN : 0632057599
Johann, L., Dinace, C., & Shapiro, S. A. (2016). Scaling of seismicity induced by nonlinear fluid-rock interaction after an injection stop. Journal of Geophysical Research: Solid Earth, 121, 8154–8174. https://doi.org/10.1002/2016JB012949
Kauffman, L., & Rousseveu, P. (1987). Clustering by means of medoids. North-Holland
Kauffman, L., & Rousseveu, P. J. (2009). Finding groups in data: An introduction to cluster analysis (Vol. 344). Hoboken, NJ: John Wiley & Sons.
Kazor, K., & Hering, A. S. (2015). Assessing the performance of model-based clustering methods in multivariate time series with application to identifying regional wind regimes. Journal of Agricultural, Biological, and Environmental Statistics, 20(2), 192–217. https://doi.org/10.1007/s13253-015-0205-8
Kim, K.-H., Yun, S.-T., Park, S.-S., Joo, Y., & Kim, T.-S. (2014). Model-based clustering of hydrochemical data to demarcate natural versus human impacts on bedrock groundwater quality in rural areas, South Korea. *Journal of Hydrology*, 519, 626–636. https://doi.org/10.1016/j.jhydrol.2014.07.055

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). USA: University of California Press.

Maxwell, S. C., Mack, M., Zhang, F., Chorney, D., Goodfellow, S. D., & Grob, M. (2015). Differentiating wet and dry microseismic events induced during hydraulic fracturing. In Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2015-2154344

Maxwell, S. C., Waltman, C., Warpinski, N. R., Mayerhofer, M. J., & Boroumand, N. (2009). Imaging seismic deformation induced by hydraulic fracture complexity. *SPE Reservoir Evaluation & Evaluation*, 12(1), 48–52. https://doi.org/10.2118/102801-pa

Maxwell, S. C., Zhang, F., & Damjanac, B. (2015). Geomechanical modeling of induced seismicity resulting from hydraulic fracturing. *The Leading Edge*, 34(6), 678–683. https://doi.org/10.1190/tle3406678.1

Moriya, H., Niitsuma, H., & Baria, R. (2005). Multiplet analysis for estimation of critical pore pressure of fractures inside geothermal reservoir. In Proceedings of the World Geothermal Congress 2005.

Munson, E. O. (2015). Reservoir characterization of the Duvernay Formation, Alberta: A pore to basin-scale investigation. Thesis. University of British Columbia. https://doi.org/10.14288/1.0166722.

Nelson, R. (2001). *Geologic analysis of naturally fractured reservoirs*. Boston: Gulf Professional Publishing.

Nordgren, P. (1972). Propagation of a vertical hydraulic fracture. *Society of Petroleum Engineers Journal*, 25(3), 306–314.

Pearson, K. (1897). *Mathematical contributions to the theory of evolution*—On a form of spurious correlation which may arise when indices stated in m. Baitre are used in the measurement of organs. *Proceedings of the Royal Society of London, Series A, Containing Papers of a Mathematical and Physical Character*, 58(405–407), 489–498.

Pearson, K. (1901). *LIII. On lines and planes of closest fit to systems of points in space*. The *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. https://doi.org/10.1080/14786440109462720

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. 2013

Rahman, M. M., & Govindaraju, Z. (1997). A modification of the test of Shapiro and Wilk for normality. *Journal of Applied Statistics*, 24(2), 219–236. https://doi.org/10.1080/026647697238282

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.

Rutledge, J. T., & Phillips, W. S. (2003). Hydraulic stimulation of natural fractures as revealed by induced microearthquakes. *Carthage Cotton Valley gas field, east Texas hydraulic stimulation of natural fractures*. *Geophysics*, 68(2), 441–452. https://doi.org/10.1190/1.1567214

Sander, J., Ester, M., Kriegel, H. P., Xu, X. (1998). *Data Mining and Knowledge Discovery*. 2, 169. https://doi.org/10.1023/A:1009745219419

Schultz, R., Wang, R., Gu, Y.-J., Haug, K., & Atkinson, G. (2017). A seismological overview of the induced earthquakes in the Duvernay play near Fox Creek, Alberta. *Journal of Geophysical Research: Solid Earth*, 122, 492–505. https://doi.org/10.1002/2016JB013570

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. https://doi.org/10.1214/aos/1176344316

Shapiro, A. S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52(3), 591–611. https://doi.org/10.2307/233709

Shapiro, S. A. (2015). *Fluid-induced seismicity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139051132

Shapiro, S. A., Dinske, C., & Rothert, E. (2006). Hydraulic fracturing controlled dynamics of microseismic clouds. *Geophysical Research Letters*, 33, L14312. https://doi.org/10.1029/2006GL026365

Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2015). Earthquakes induced by hydraulic fracturing in Poland Township, Ohio. *Bulletin of the Seismological Society of America*, 105(1), 189–197. https://doi.org/10.1785/0120140168

Smith, M. B., & Montgomery, C. T. (2015). *Hydraulic fracturing* (1st ed.). Boca Raton, Florida: CRC Press. https://doi.org/10.1201/b16287

Smith, C. A. (1964). The proof and measurement of association between two things. *The American Journal of Psychology*, 77(1), 72–101. https://doi.org/10.2307/1412159

Tezuka, K., & Nitsuma, H. (2000). Stress estimated using microseismic clusters and its relationship to the fracture system of the Hijiori hot dry rock reservoir. *Engineering Geology*, 561–2, 47–62. https://doi.org/10.1016/S0013-7952(99)00133-7

The Royal Society (2017). *Machine learning: the power and promise of computers that learn by example*. London, UK: The Royal Society.

Valko, P., & Economides, M. J. (1995). *Hydraulic fracturing mechanics*. John Wiley & Sons, Hoboken, NJ. ISBN: 0471956643

Vasudevan, K., Eaton, D. W., & Davidsen, J. (2010). Intraplate seismicity in Canada: A graph theoretic approach to data analysis and interpretation. *Nonlinear Processes in Geophysics*, 17(5), 513–527. https://doi.org/10.5194/npg-17-513-2010

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Warpinski, N. R. (2013). Understanding hydraulic fracture growth, effectiveness, and safety through microseismic monitoring. *International Society for Rock Mechanics and Rock Engineering*. https://doi.org/10.5194/isrm-2013-26

Wei, T., & Simko, V. (2013). corplot: Visualization of a correlation matrix. *R Package Version 0.73, 230(231), 11.

Wiener, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4), 459–469. https://doi.org/10.1785/019990114

Williamson, H. S. (1999). Accuracy prediction for directional MWD. Society of Petroleum Engineers. https://doi.org/10.2118/56702-MS