GEOPHYSICS, VOL. 75, NO. 3 (MAY-JUNE 2010); P. L61–L70, 5 FIGS., 1 TABLE.
10.1190/1.3429997

Cluster analysis of Euler deconvolution solutions: New filtering techniques and geologic strike determination

Hernan Ugalde[1] and William A. Morris[1]

ABSTRACT

Euler deconvolution often presents the problem of filtering coherent solutions from uncorrelated ones. We have applied clustering and kernel density distribution techniques to a Euler-generated data set. First a kernel density distribution algorithm filters uncorrelated Euler solutions from those consistently located near an anomalous magnetic-gravimetric source. Then a fuzzy *c*-means clustering algorithm is applied to the filtered data set. The computation of cluster centers reduces the size of the data set considerably, yet maintains its statistical consistency. Finally, the computation of eigenvectors and eigenvalues on the cluster centers yields an estimate of the geologic strike of the anomalous sources responsible for the observed geophysical anomalies. Therefore, we can obtain an improved strike and depth estimation of the magnetic sources. Although the algorithm can filter and cluster any Euler data set, we recommend obtaining the best solutions possible before any clustering. Hence, we have used a hybrid 3D extended Euler and 3D Werner deconvolution algorithm. We have developed synthetic and real examples from the Bathurst Mining Camp (New Brunswick, Canada). The output of this algorithm can be used as an input to any 3D geologic-modeling package.

INTRODUCTION

Euler deconvolution has become a very popular tool as an aid to interpreting profile or gridded magnetic data (Reid et al., 1990). It provides automatic estimates of anomalous source locations and their depths. In its usual standard 3D form, it requires a priori knowledge of the rate of decay of the field (known as the structural index). This method can lead to wrong results if the geologic knowledge of the associated sources of the observed potential field anomalies is limited (Reid et al., 1990).

Low signal-to-noise ratio, violations of the homogeneity condition involved in the basic Euler equation, and superposition of sources all cause Euler solutions to form broad clouds instead of dense clusters, making it difficult to outline the boundaries of the associated causative sources. Traditional discrimination techniques used to remove spurious solutions are based on a combination of the following criteria (Fitzgerald et al., 2004): a threshold of the ratio of estimated depth to the standard deviation of the depth (Thompson, 1982); solutions falling outside the x- and y-coordinates of the present data window; a threshold of signal strength of the anomaly; spatial binning of solutions; and low-pass filtering of the data to constrain the frequency content (i.e., rejecting high-frequency features associated with shallow sources) before running the actual deconvolution. However, these discrimination techniques cannot evaluate the overall quality of the solutions or determine how dense/compact the clusters are.

Silva and Barbosa (2003) derive estimators for the horizontal and vertical source positions as a function of the *x*-, *y*- and *z*-gradients of a magnetic anomaly within a data window. In that way, they can constrain the region where good solutions will be generated and therefore decrease the overall dispersion. Gerovska and Arauzo-Bravo (2003) present a different approach for decoupling the coordinates of the magnetic sources from their associated structural index by using a differential similarity transform (DST). They also cluster the solution database and apply statistics to filter out spurious *x*-, *y*-, and *z*-coordinates. Other innovative techniques for filtering and classifying the Euler output database involve the use of artificial intelligence (Mikhailov et al., 2003).

Extensions of Euler deconvolution, such as the 2D extended Euler deconvolution (Mushayandebvu et al., 2001) or combined analytic signal with Euler (AN-EUL) (Salem and Ravat, 2003), rely on either the use of simplified models (thin-dike, vertical contacts) (Mushayandebvu et al., 2001; Cooper, 2006), which limits the applicability of the method to only specific cases; the combined solution of Euler deconvolution by means of using analytic signal peaks (Salem and Ravat, 2003); or the application of Euler deconvolution to the analytic signal, thereby solving for the source location and structural index (Keating and Pilkington, 2004; Florio et al., 2006).

Manuscript received by the Editor 12 March 2009; revised manuscript received 20 December 2009; published online 9 June 2010.
[1]McMaster University, School of Geography and Earth Sciences, Hamilton, Ontario, Canada. E-mail: ugaldeh@mcmaster.ca; morriswa@mcmaster.ca.

Interpretation of the analytic signal amplitude (ASA) (Nabighian, 1972; Roest et al., 1992) instead of regular total magnetic intensity (TMI) has many advantages, such as enhancing the gradients and concentrating the gradients of all three directions in one function. However, like any gradient-based method, it will suffer from noise enhancement unless care is taken when computing the derivatives. The AN-EUL method works well in cases when there is sufficient signal for the ASA to map edges or contacts effectively. However, because part of the algorithm involves finding the peaks of the ASA and then solving the Euler equation on those peaks, it has problems in areas like deep sedimentary basins, where the sources are deeply buried and only produce small amplitude ASA peaks.

In this article, we focus on the classification and filtering of Euler solutions and the subsequent extraction of geologic information (strike, depth) that can be used by other geologic-modeling packages. Although here we test the algorithm with a hybrid 3D extended Euler and 3D Werner deconvolution data set (Nabighian and Hansen, 2001; Reid and Fitzgerald, 2005), the selection and filtering of solutions and the extraction of structural geologic information is an important matter regardless of the variation of Euler deconvolution being used. This work focuses on the use of kernel density estimation and clustering analysis techniques. This method can therefore be applied to any Euler solution data set, regardless of the computational method.

METHODOLOGY

This work focuses on the analysis of Euler deconvolution solutions instead of Euler deconvolution computation. Therefore our method can start with the output of any Euler deconvolution algorithm, and then we classify the Euler solution data set. Classification of the data set is accomplished according to similarities among the data. However, to obtain good results, we recommend using the best Euler algorithm for the first stage, and therefore we use a hybrid 3D extended Euler and 3D Werner algorithm (Nabighian and Hansen, 2001; Reid and Fitzgerald, 2005).

The 3D equation for Euler deconvolution is originally given by Reid et al. (1990) and rewritten by Nabighian and Hansen (2001) as

$$(x-x_0)\frac{\partial T}{\partial x}+(y-y_0)\frac{\partial T}{\partial y}+(z-z_0)\frac{\partial T}{\partial z}+N(T-B)=\alpha, \tag{1}$$

where (x_0,y_0,z_0) is the position of a source whose total magnetic field T is measured at a point (x,y,z); B and N are the regional magnetic field and structural index, respectively, and α is a constant that normally vanishes except for $N=0$.

Mushayandebvu et al. (2001) derive the extended Euler method from the observation that potential fields are invariant under rotations. Applying Hilbert transforms in the local east (x) and north (y), giving $H_x(T)$ and $H_y(T)$, respectively, leads to two new equations (Fitzgerald et al., 2004):

$$(x-x_0)\frac{\partial H_x(T)}{\partial x}+(y-y_0)\frac{\partial H_x(T)}{\partial y}+(z-z_0)\frac{\partial H_x(T)}{\partial z}+NH_x(T)=\beta_x, \tag{2}$$

$$(x-x_0)\frac{\partial H_y(T)}{\partial x}+(y-y_0)\frac{\partial H_y(T)}{\partial y}+(z-z_0)\frac{\partial H_y(T)}{\partial z}+NH_y(T)=\beta_y, \tag{3}$$

where β_x and β_y are constants that also vanish except for $N=0$.

This general result, that if a solution of Laplace's equation satisfies a Euler equation with index N, so do its generalized Hilbert transforms (Nabighian and Hansen, 2001), allows the direct solution for depth with reduced uncertainty and with no need to assume any fixed value of the structural index (Reid and Fitzgerald, 2005).

Because the 3D Euler deconvolution method estimates the horizontal and vertical source coordinates for each position of the moving data window spanning the gridded data, Euler solutions form broad clouds. The main idea of our classification approach is the hypothesis that passing a moving data window over an area of a real geologic source and its associated geophysical anomaly should result in a population of similar Euler solutions with a mean (or median, to minimize the effect of outliers) that approximates the optimum solution. The identification of these zones is achieved through an approach of combined cluster analysis and kernel density estimation. Once the best possible Euler solution data set has been extracted from Euler deconvolution (in this case, 3D extended Euler and 3D Werner deconvolutions combined), the solution-classification and filtering algorithm is divided into three parts: (1) kernel density estimation to reject uncorrelated solutions, (2) fuzzy c-means clustering (FCM) of the filtered solutions (see Appendix A), and (3) eigenvector computation on the clustered (filtered) data set to derive the main geologic strike of the mapped structures.

Kernel density estimation

Kernel density estimation is used to reject solutions that are not part of a dense cluster and that are sparsely located, and do not follow a consistent geologic trend and can therefore be classified as undesired. A trivariate kernel density estimation is used to find the likelihood of a point $\mathbf{x}_0=(x_0,y_0,z_0)$ to be in the vicinity of a cluster center $\mathbf{x}_c=(x,y,z)$.

The trivariate kernel density distribution, with kernel K and window width h of a sample of a random variable $(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$, is given by

$$f_h(\mathbf{x})=\frac{1}{nh^3}\sum_{i=1}^{n}K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right), \tag{4}$$

where $\mathbf{x},\mathbf{x}_i\in\mathbb{R}^3$, n is the sample size (in this case, the number of Euler solutions being analyzed), and K is a radially symmetrical trivariate density distribution, defined by

$$K(\mathbf{x})=(2\pi)^{-3/2}\exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x}\right). \tag{5}$$

As h tends to zero, its effect on the kernel density distribution will be a sum of Dirac delta functions at the observation points, thus emphasizing noisy features in the data, whereas as h becomes large, all detail in the data is obscured because of oversmoothing (Silverman, 1986). Therefore, an optimum value of h must be chosen between those extremes.

Given an n-point sample $\mathbf{x}$ of a function, with standard deviation σ, the normal scale rule allows the computation of h as

$$h = \frac{1.6\sigma}{\sqrt[5]{n}}. \quad (6)$$

This result is valid for normally distributed data sets. For nonnormal, multimodal data, the normal scale rule will oversmooth the data. In practice, given a Euler solutions data set of size n and standard deviation σ, equation 6 is used to compute the window size h. Equations 4 and 5 are then used to compute the trivariate kernel density distribution of the data set.

The kernel density estimation filters uncorrelated solutions from those that do form dense clouds. To do that, a threshold is chosen by inspection to ensure that most of the uncorrelated solutions are rejected and most of the coherent data are kept. The next step is applied on the filtered data set.

Fuzzy *c*-means clustering

Once the selected Euler solution data set has been classified in terms of coherent solutions versus undesired solutions, an FCM algorithm is applied to obtain the centers (x,y,z) of c clusters within the filtered data set. These clusters represent the main geologic structures being mapped. The major drawback of the algorithm is that the number of clusters c needs to be defined before the actual computation (Balasko et al., 2005). One alternative is to do the clustering for different values of c and then compute validity indexes for every partition to assess the goodness of each. Some common indexes used for this matter are the partition index (SCI) (Balasko et al., 2005) and the Xie and Beni index (XBI) (Xie and Beni, 1991). The expressions for both indexes are given in Appendix A.

The SCI is the ratio of the sum of compactness and separation of the clusters. It is useful when comparing different partitions having the same number of clusters. A lower value of SCI indicates a better partition. The XBI is a normalized index between 0 and 1, and the c for which the XBI is minimum indicates the optimum number of clusters. In our case, both XBI and SCI gave consistently similar results, so we were able to use them jointly. In case of a large disparity between them, we recommend plotting the cluster locations for both cases and doing a close inspection, and then deciding which result is more meaningful from a geologic perspective.

Determination of geologic strike of sources mapped

The interpreter must define a distance threshold around each cluster to keep only those points within that defined distance from each cluster. A common threshold is being used for all clusters, but a threshold could be defined specifically for each cluster, although that would require more user input and some degree of automation would be lost. In our case, this common threshold was chosen by inspection; however, its value could be determined automatically by some norm, such as a percentage of the average distance between data points and cluster centers.

Given a set of c clusters found in the previous stage, we define for each cluster N_j as the number of points preserved on the jth, cluster c_j, $j = 1,\ldots,c$. We build the partition matrix $\mathbf{D}_{ikA}$ according to equation A-10. Eigenvectors and eigenvalues are computed for each of these 3×3 matrices. The eigenvalues provide a means of ranking the goodness of the clusters even further, and the eigenvectors give the orientation (strike and dip) of the cluster center (x,y,z) that estimates the location of the anomalous sources. Eigenvectors on elongated structures can define the strike of the geologic feature of interest.

Although this technique can calculate and provide the actual 3D orientation of any 3D structure defined by a data set of points in 3D space, in real terms the Euler deconvolution solutions cannot map the dip of the associated anomalous geologic structures. Therefore, we will focus the analysis of the data in terms of geologic strike. Only the eigenvector associated to the major eigenvalue was considered for each cluster because this is more representative for the 3D orientation of its associated points. A further extension of this approach would be to consider the eigenvectors associated with the two major eigenvalues, thus determining the shape and orientation of the associated source. In the case of one larger eigenvalue, the source would be elongated; two similar eigenvalues would yield a 2D source with similar dimensions on both major orientations, along strike and perpendicular to it. Mushayandebvu et al. (2004) describe a similar approach for the discrimination between 2D and 3D sources, based on the finding of eigenvalues very close to zero.

SYNTHETIC MODEL RESULTS

The algorithm was tested with both synthetic and real data. Because one of the main goals was to test the ability of the algorithm to solve for the strike of geologic structures, the synthetic model consisted of two highly elongated bodies (Figure 1a). Both bodies were 500 m thick and had a depth extent of 2000 m. The north-south body was 4 km long; the east-west one was 8 km long. The depth to top was 50 m in both cases. The bodies dip at 45° to the east (north-south body) and 45° to the south (east-west body). The ambient field parameters were total intensity 60,000 nT, inclination 90°, and declination 0°. No remanent magnetization was added. The magnetic susceptibility of both bodies was 0.0035 structural index. A total magnetic intensity (TMI) grid was calculated with north-south lines spaced at 500 m, and at a common observation level of 0 m (Figure 1b). Along the lines, spacing was 10 m. The forward model was calculated on profiles and then gridded at a 100-m grid cell size by means of a standard minimum curvature algorithm. Figure 1c shows a north–south profile through the center of the area. Both the grid and profile views of the TMI model show the expected asymmetry of the magnetic anomalies due to the dip of the bodies.

A hybrid 3D extended Euler and 3D Werner deconvolution algorithm was applied to obtain the depth-to-source solutions (Nabighian and Hansen, 2001; Reid and Fitzgerald, 2005). To concentrate the solutions near the top of the bodies, a window of 500×500 m was used, and as explained in the previous section, the algorithm solved for the structural index of the sources too, so no a priori information on the structural index was required. The algorithm generated approximately 12,300 solutions over the area of 20×20 km. After some standard masking based on the structural index (SI <1.5 to preserve sources associated with contacts and thin dikes) and error constraints, the solution database was reduced to approximately 4500 solutions (blue crosses in Figure 2a). Although the Euler solution data set is still quite large, it can map the top of both bodies very well.

Subsequently, we applied the kernel density algorithm. We use a Gaussian kernel (equation 5) and the normal scale rule (equation 6)

to determine the smoothing factor h. Figure 2a shows the results of the kernel density classification (yellow volume), superimposed on the original Euler solution data set (blue). A 70% probability was used as the density cutoff. This value ensured that most of the uncorrelated solutions were removed, without losing useful ones. The application was able to resolve the areas of higher probability from those where the Euler solutions were uncorrelated. Only the points within the yellow volume were kept as valid points. The FCM clustering was applied to a different number of clusters from 2 to 30. The XBI and SCI indexes were computed to determine the optimum number of clusters present on the data set, corresponding to the minimum value of each index. The XBI showed a minimum at $c = 21$, and the SCI had its minimum at $c = 20$. Both numbers were tried and yielded almost identical results on cluster locations. A value of $c = 20$ was kept as the number of clusters for the next calculations.

Finally, the filtered Euler solution data set was clustered via the FCM algorithm, and the eigenvectors were computed to determine the strike and of the observed sources. Figure 2b and c shows the location of the 20 clusters, again plotted with the entire original data set for reference. Table 1 summarizes, for each cluster, the estimated strike from the eigenvector associated with the major eigenvalue. The results are very consistent, giving a mean strike of 87° for the east-west body (true strike 90°) and 5° for the north-south one (true strike 0°).

The result of kernel density combined with FCM clustering did indeed filter the uncorrelated solutions out of the original Euler solution data set, and it was successful in mapping the geologic strike of 2D dipping sources. The only areas where the estimated strikes showed some dispersion were the corners of the bodies, which is exactly where a discontinuity exists from 2D to 3D and therefore the Euler system is not stable (Mushayandebvu et al., 2004).

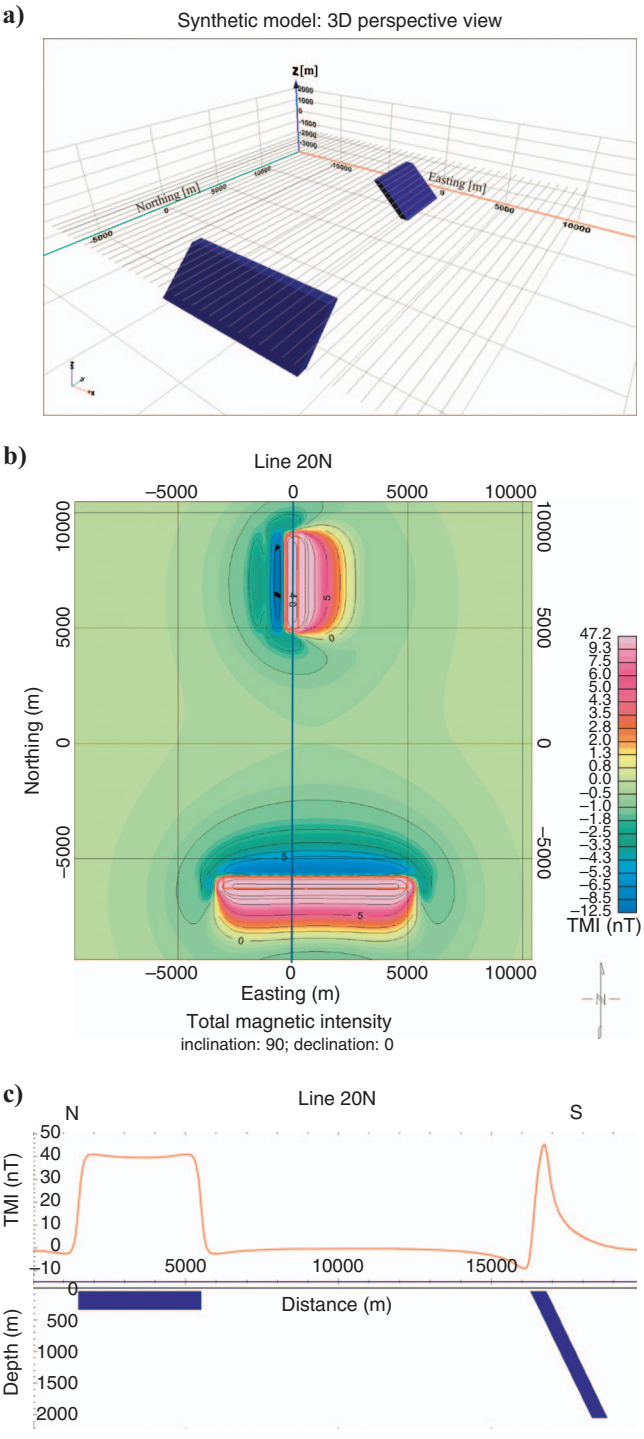


Figure 1. Synthetic model. (a) A 3D perspective view of the causative bodies; view from the southeast. (b) Total magnetic intensity caused by the two dipping bodies. See text for details on the model parameters. (c) North-south cross section along the central part of the grid (green line on Part b).

APPLICATION: BATHURST MINING CAMP, NEW BRUNSWICK, CANADA

Finally, we applied the algorithm on a real data set at the Bathurst Mining Camp (BMC) in New Brunswick, Canada (Figure 3). The BMC is one of the most important base-metal mining districts in Canada. The camp hosts numerous sediment- and volcanic-hosted massive sulfide deposits and occurrences. However, most major deposits were discovered in the 1950s and 1960s; therefore new discoveries are critical to avoid a drastic decline in production over the next years. Outcrop in the area is scarce; therefore any structural information that can be derived from airborne geophysical data is invaluable. Figure 3 shows the regional geology of the BMC (after Galley et al., 2007).

The BMC is primarily composed of Cambrian and Ordovician sedimentary and volcanic rock units, which have been subject to a complex history of multigenerational folding, thrusting, and faulting (Van Staal et al., 2003). An airborne geophysical survey was flown in 1996 to aid in the geologic mapping and exploration of the camp. The line spacing was 200 m, and the magnetic sensor was towed at 60 m above the ground. Flight lines were oriented perpendicular to the main geologic strike in each of the four subblocks in which the entire camp was divided (Geological Survey of Canada, 1996).

Figure 4a shows the total magnetic field reduced to the pole total magnetic intensity (RTP-TMI) data over a window of the BMC where we tested Euler deconvolution and the clustering algorithm. The magnetic data show the contrast between weakly magnetic felsic volcanic units and basalts, gabbros, and strongly magnetic mafic volcanic and intrusive units (Keating et al., 2003). The locations of two mineral deposits in the area (Caribou and McMaster) are shown for reference. Figure 4b shows a 3D perspective view of the geologic map draped over the magnetic data. The correlation between the data

sets is good, although some areas apparently do not have any associated magnetic signature, like the folded units on the southeast part of the map. Any additional structural information derived from semiautomatic interpretation of geophysical data can assist in improving geologic mapping in areas of scarce outcrop such as the BMC.

As with the synthetic data set, a hybrid 3D extended Euler and 3D Werner deconvolution was applied to the RTP data shown in Figure 4a. The algorithm produced approximately 27,000 (x,y,z) solutions with depths between 0 and 1000 m (a window size of 1000 m was used). After the usual removal of spurious solutions with high spatial (dXY) and depth uncertainties (dZ) and higher structural index so as to preserve solutions associated with contacts, the Euler solution data set was reduced to approximately 4000 solutions, which show good correlation with the main geologic contacts (Figure 5a-c). The kernel density routine was applied first to reduce the number of solutions even further (Figure 5a). The final results required a lower probability density cutoff than on the synthetic model (60% here, as compared with 70% on the synthetic model). This was due to the different nature of the data: many small clusters with a few outliers, compared with a couple of large clusters with many outliers on the synthetic model. Subsequently, the FCM clustering algorithm was applied to this data set. After running it with different cluster numbers and determining the minimum on both XBI and SCI indexes, an optimum number of 123 clusters were determined. Cluster locations are shown on Figure 5b, and the structural information derived from them is shown on Figure 5c.

The strike information derived from this algorithm follows the folded structures in the BMC quite closely, and it should be of good use for any subsequent geologic-modeling attempts in the area.

DISCUSSION

The algorithm presented here is far from an automatic recipe for structural mapping from geophysical data. There are many parameters to be determined during the calculation, and good results depend on a proper selection of them. Nonetheless, the results of the combined kernel density and FCM clustering algorithm on a good-quality Euler data set are quite satisfactory and should be of good use as input for geologic modeling and structural analysis.

Undoubtedly, the algorithm selected for the Euler deconvolution is critical for the success of the final result (geologic strike and depth estimations). Although the methodology presented here will filter data even if the Euler solution data set is not well constrained, the strike and depth information obtained will not be meaningful unless the data are well correlated and follow actual geologic structures. The hybrid 3D extended Euler and 3D Werner method used here gave excellent results in terms of solution coherency; therefore the strike information was authentic in both the synthetic and real data examples.

The kernel density estimation does a good job of filtering undesired Euler solutions. However, it seems to have some difficulties when the data are sparsely located, as on the BMC example shown here. The algorithm works by computing the kernel density estimation and then filtering the data based on a percentage cutoff or mini-

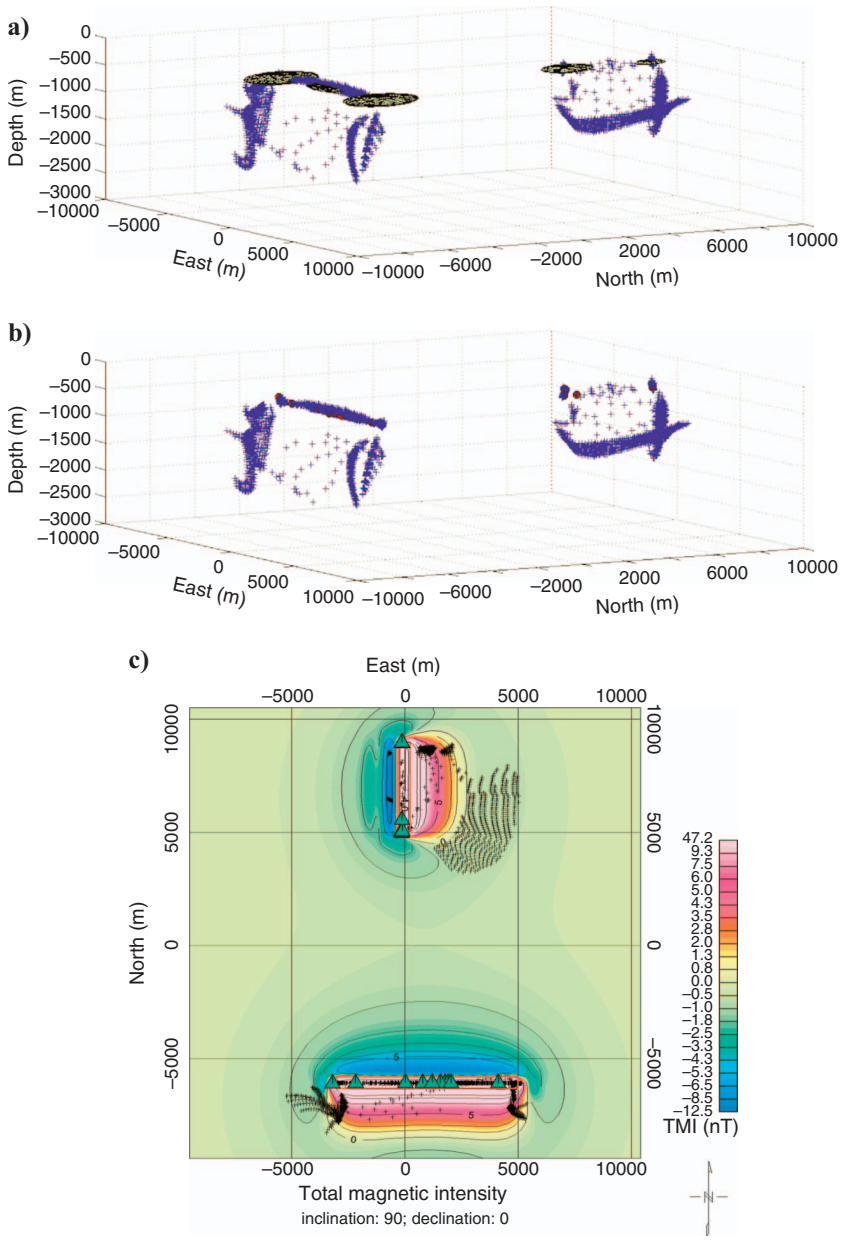


Figure 2. Application of the kernel density and FCM algorithm to Euler deconvolution solutions on the synthetic model of Figure 1. (a) A 3D perspective view of the Euler deconvolution solution data set (blue crosses), and the kernel density volumes (yellow). For the clustering, only the points within the yellow volumes are selected. (b) A 3D perspective view of the cluster locations (red circles). (c) Plan view of the original magnetic data set with the original Euler deconvolution solution data set (blue dots), and the cluster locations (green triangles). See text for details on the calculation parameters.

mum probability. Thus, it tends to generate large closed surfaces at that probability cutoff. For the case of distributions where the correlated data cluster in only a few areas with noncorrelated points around (e.g., example in Figure 2a), enclosing the higher probability points in a large surface does work well. For the case in which the division between correlated and uncorrelated data is more diffuse because of the presence of multiple small clusters with some uncorrelated data around, the algorithm requires more trials to find the right probability cutoff.One option not tried here is to divide the data into smaller windows and apply the kernel density estimation and cutoff on those smaller windows.

The FCM clustering algorithm is very stable. Both applications (synthetic and real data) gave very good results in terms of partitioning the data into proper clusters and recognizing the major clouds from the sparsely located solutions.

The addition of the eigenvector computation and its associated structural information is undoubtedly a major step forward toward linking geophysical data processing with geologic modeling and interpretation. Although in cases like the BMC, where the visual inspection of the magnetic anomalies does indeed give the interpreter agood idea of the geologic strike of the main structures, a systematic and unbiased approach like this one goes a step further, automating the process and providing a numerical table of coordinates and strike, which can subsequently be used as input for other geologic-modeling packages.

Although the methodology was applied mostly to derive the strike directions and depth of anomalous bodies, it undoubtedly can be applied to other problems such as model construction (instead of inversion for physical properties) and depth to basement, which can be solved through the analysis of the depth to top of anomalous bodies through Euler deconvolution and the filtering and classification scheme presented here.

Table 1. Cluster locations, major eigenvalue (EigVal), associated eigenvector (Eig), and computed strike (Str) information for the results on the synthetic data set. For consistency in the data analysis, the table is divided into two segments associated to the east-west body (top 10 solutions) and north-south body (bottom 6 solutions).

X (m)	*Y* (m)	*Z* (m)	Eig *X*	Eig *Y*	Eig *Z*	EigVal	Str (°)
	East-west body						
4563.31	− 7269.95	− 1646.02	0.8532	0.0920	0.9843	0.03	83.85
4537.23	− 7265.22	− 1728.76	0.8798	0.0825	0.9862	0.03	84.64
− 2585.29	− 7229.94	− 1247.43	0.9765	0.0993	0.9945	0.08	84.19
4573.30	− 7198.62	− 1552.45	− 0.9536	0.0907	0.9910	0.04	84.57
4560.67	− 6981.53	− 1318.63	− 0.9754	0.0253	0.9859	0.04	88.51
2752.97	− 6264.22	− 119.16	0.9997	− 0.0249	− 0.0016	0.32	88.57
− 3169.00	− 6191.54	− 144.99	0.9858	− 0.0670	0.1537	0.33	86.11
5063.47	− 6116.29	− 452.93	− 0.8502	0.0406	0.5248	0.13	87.27
− 3126.92	− 6096.90	− 502.10	0.6442	0.0141	0.7648	0.14	88.75
758.31	− 6042.17	− 283.40	0.9861	0.0061	− 0.1660	0.19	89.65
5047.44	− 6037.72	− 161.01	− 0.9961	0.0149	0.0873	0.26	89.14
− 463.42	− 6030.20	− 288.26	0.9819	− 0.0355	0.1859	0.19	87.93
3520.49	− 6021.09	− 303.21	0.9979	− 0.0553	0.0348	0.18	86.83
2053.45	− 6016.91	− 301.29	0.9790	0.0049	− 0.2038	0.21	89.71
− 1779.85	− 5976.17	− 288.98	0.9738	− 0.0209	0.2266	0.22	88.77
						Mean	87.23
						Std	2.07
	North-south body						
1109.48	5100.40	− 676.77	0.0373	− 0.5058	− 0.9937	0.17	4.22
732.89	5224.03	− 168.01	− 0.1379	0.8216	− 0.5180	0.19	9.53
1575.17	5318.20	− 884.04	0.0126	− 0.3523	− 0.9932	0.21	2.05
− 247.93	7303.77	− 179.06	0.0513	0.8338	− 0.2105	0.31	3.52
987.74	8900.00	− 522.25	− 0.0888	0.8453	0.9925	0.46	6.00
						Mean	5.06
						Std	2.87

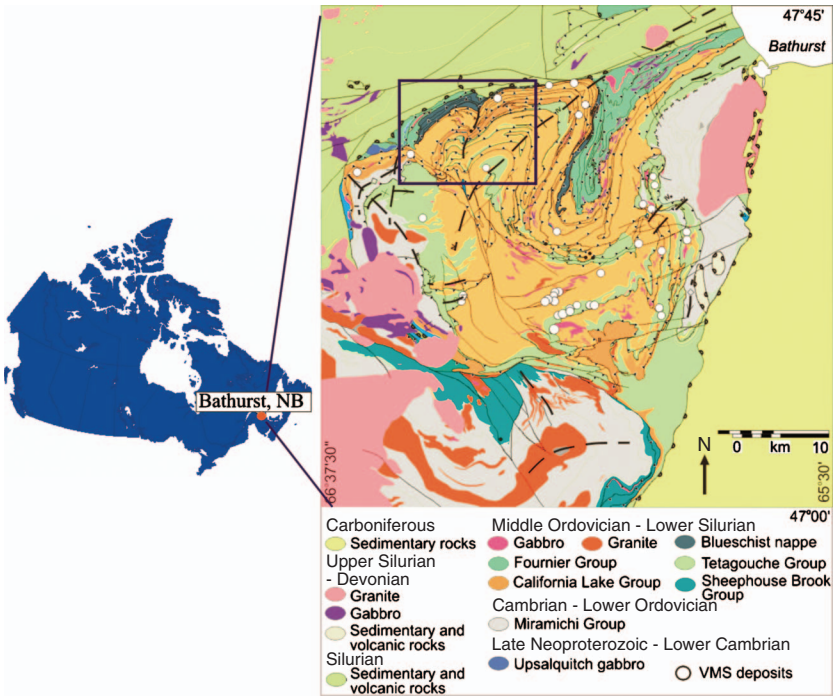
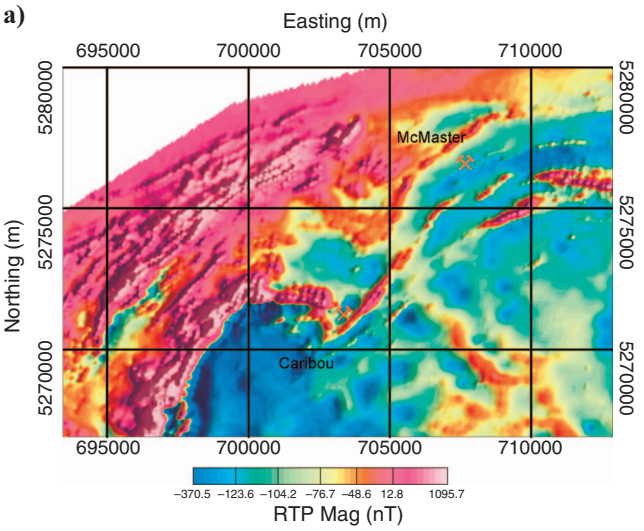


Figure 3. Location map of the Bathurst Mining Camp, New Brunswick, Canada (left) and regional geology (right); VMS indicates volcanogenic massive sulfide. The blue rectangle on the geologic map marks the approximate location of area shown on Figures 4 and 5. Regional geology from Galley et al. (2007).



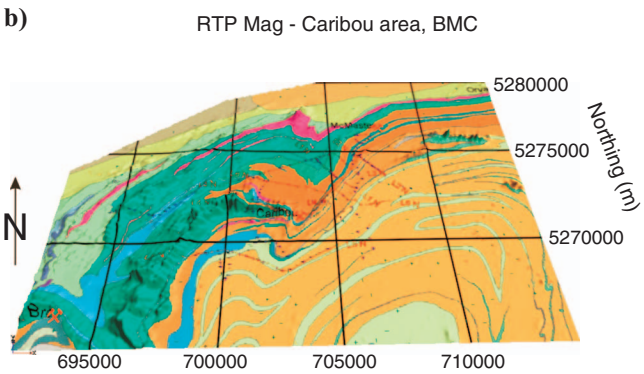


Figure 4. (a) Pole-reduced total magnetic intensity map over the area of study. The locations of two deposits in the area (Caribou and McMaster) are shown for reference. (b) A 3D composite built with the regional geologic map draped over the RTP map from (a).

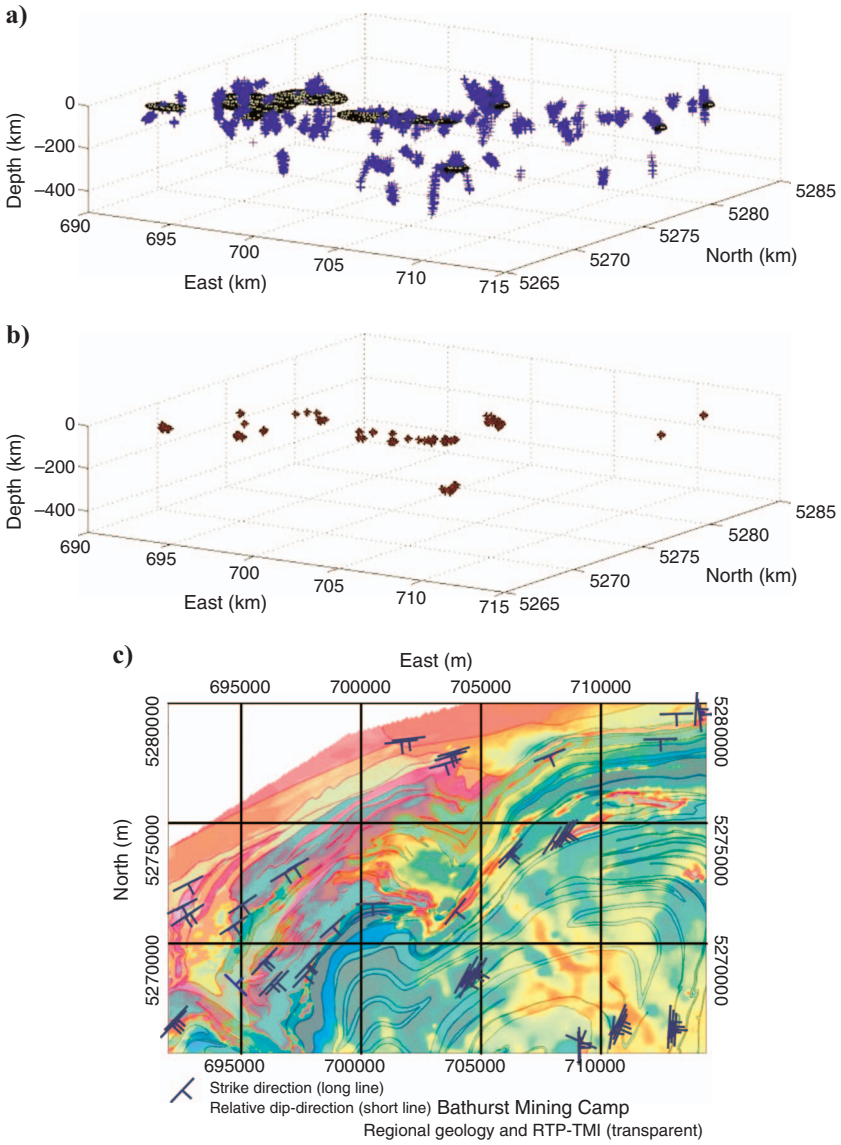


Figure 5. Application of the kernel density and FCM clustering algorithm to the BMC data set. (a) A 3D perspective view of the kernel density (yellow), with the original Euler deconvolution solutions (blue crosses) for reference. (b) Cluster locations (red stars), over the kernel density filtered data set. (c) Regional geology of the BMC (after Galley et al., 2007) and RTP-TMI data from (a), 50% transparency. Strike estimations from the primary eigenvectors of each cluster are shown as blue lines.

CONCLUSIONS

We present a new algorithm for the filtering of Euler solutions. The method not only can discriminate between "good" and "bad" solutions, but also provide the strike and average depth of the geologic sources, which can then be used as input for 3D geologic-modeling algorithms. The algorithm was tested on one synthetic data set and one real data set at the Bathurst Mining Camp (BMC) in New Brunswick, Canada. At the BMC, the kernel density distribution and clustering method presented here was able to add some structural information over an area of scarce outcrop.

Further work involves the refinement of the kernel density and clustering algorithms to determine optimum threshold values for the kernel density and number of clusters. In its present state, the algorithm requires the FCM clustering routine to be run with different cluster numbers to determine the optimum solution set. This is computationally inefficient and can be considerably slow for large data sets. The minimization of anomaly interference effects through curvature-based filtering of magnetic/gravity sources prior to the application of Euler deconvolution is an area that needs further development as well.

ACKNOWLEDGMENTS

Financial support for this study was provided by an INTERACT grant through the Ontario Centres of Excellence (OCE): Crestech, Paterson, Grant & Watson Limited (PGW) and DGI Geosciences. Data in the Bathurst Mining Camp were provided by the Geological Survey of Canada (GSC). Constructive reviews on earlier versions of the manuscript by the associate editors M. Pilkington and V. Barbosa, the reviewers S. Goussev and D. Fitzgerald, and two anonymous reviewers improved the article considerably.

APPENDIX A

FUZZY *C*-MEANS CLUSTERING

General definitions

Assume that we have a data set representing observations of n variables. Each observation consists of n measured variables grouped into an n-dimensional vector $\mathbf{x}_k \equiv (x_{k1}, x_{k2}, \ldots, x_{kn})^T$. Let $\mathbf{X}$ be an $N \times n$ data matrix whose rows correspond to the N observations and whose columns correspond to the n variables defining each solution:

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nn} \end{pmatrix}. \qquad \text{(A-1)}$$

Depending on the objective of the clustering, various definitions of a cluster can be formulated. Generally, one might accept that a cluster is a group of objects that are more similar to one another than to members of other clusters. In metric spaces, the term "similarity" is often defined by means of a distance norm.

Clusters are subsets of the main data set. In general terms, two clustering methods can be defined, hard or fuzzy. Hard clustering requires that an object either does or does not belong to a cluster. For a given data set $\mathbf{X}$, hard clustering means partitioning it into c mutually exclusive subsets of $\mathbf{X}$ (clusters). More formally, a hard partition can be defined as a family of subsets $\{\mathbf{P_i} | 1 \le i \le c \subset P(\mathbf{X})\}$ that satisfies the following properties:

$$\bigcup_{i=1}^{c} \mathbf{P_i} = \mathbf{X}, \qquad \text{(A-2)}$$

$$\mathbf{P_i} \cap \mathbf{P_j}, \quad 1 \le i \ne j \le c, \qquad \text{(A-3)}$$

$$\emptyset \subset \mathbf{P_i} \subset \mathbf{X}, \quad 1 \le i \le c. \tag{A-4}$$

In other words, all the subsets $\mathbf{P_i}$ contain all the data in $\mathbf{X}$, they are disjoint, and none of them is empty or contains all the data in $\mathbf{X}$. In terms of membership functions, where μ_i is a characteristic function of the subset $\mathbf{P_i}$, with a value of zero or one, and denoting $\mu_i(x_k)$ by μ_{ik}, we can represent hard partitions in matrix notation.

An $N \times c$ matrix $\mathbf{U} \equiv \{\mu_{ik}\}$ represents a hard partition of $\mathbf{X}$ if and only if its elements satisfy

$$\mu_{ij} \in 0,1, \quad 1 \le i \le N, \quad 1 \le k \le c, \tag{A-5}$$

$$\sum_{k=1}^{c} \mu_{ik} = 1, \quad 1 \le i \le N, \tag{A-6}$$

$$0 < \sum_{i=1}^{N} \mu_{ik} < N, \quad 1 \le k \le c. \tag{A-7}$$

Now, fuzzy partitioning can be seen as a generalization of hard partitioning, where the μ_{ik} are allowed to attain real values in the interval $[0,1]$ instead of the discrete values 0 or 1 defined in equation A-5. An $N \times c$ matrix $\mathbf{U} \equiv \{\mu_{ik}\}$ represents a fuzzy partition of $\mathbf{X}$ if and only if its elements satisfy both conditions A-6 and A-7 from above, as well as

$$\mu_{ij} \in [0,1], \quad 1 \le i \le N, \quad 1 \le k \le c. \tag{A-8}$$

Fuzzy *c*-means clustering algorithm

The fuzzy c-means clustering algorithm is based on the minimization with respect to $\mathbf{U}$ and $\mathbf{V}$ of an objective function called the c-means functional, defined as (Balasko et al., 2005)

$$\begin{aligned} J(\mathbf{X},\mathbf{U},\mathbf{V}) &= \sum_{i=1}^{c}\sum_{k=1}^{N} (\mu_{ik})^m d^2{}_{ikA} \\ &= \sum_{i=1}^{c}\sum_{k=1}^{N} (\mu_{ik})^m \|\mathbf{A}^{1/2}(\mathbf{x}_k - \mathbf{v}_i)\|^2_A, \end{aligned} \tag{A-9}$$

where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c]$, $\mathbf{v}_i \in \mathbb{R}^n$ is a matrix of cluster centers (to be determined), $\|\cdot\|$ is the Euclidian norm, $\mathbf{A}$ is an $n \times n$ diagonal matrix, and $d^2{}_{ik}$ represents the individual distances between $\mathbf{x}_k$ and $\mathbf{v}_i$. Statistically, the above equation can be seen as a measure of the total variance of $\mathbf{x}_k$ from $\mathbf{v}_i$.

The ikth element of the matrix of distances $\mathbf{D}^2$ between each data point (the vector $\mathbf{x}_k$) and every cluster center (the vector $\mathbf{v}_i$) is computed as

$$\begin{aligned} \mathbf{D}^2 \equiv \{D^2_{ik}\} &= (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A} (\mathbf{x}_k - \mathbf{v}_i), \\ & 1 \le i \le c \text{ and } 1 \le k \le N, \end{aligned} \tag{A-10}$$

where the matrix $\mathbf{A}$ can be chosen to control the shape and orientation of the clusters. If $\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is an identity nth-order matrix, the clusters will present the same shape and orientation (Balasko et al., 2005). Otherwise, we can define

$$\mathbf{A} = \begin{pmatrix} (1/\sigma_1)^2 & 0 & \ldots & 0 \\ 0 & (1/\sigma_2)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ddots & (1/\sigma_n)^2 \end{pmatrix}, \tag{A-11}$$

where σ_i is the ith variance in the direction of the coordinate axis of $\mathbf{X}$.

The minimization of $J(\mathbf{X},\mathbf{U},\mathbf{V})$ with respect to $\mathbf{U}$ and $\mathbf{V}$ is a nonlinear optimization problem that can be solved by using a variety of available methods. The fuzzy c-means algorithm iterates through the first-order conditions A-6–A-8 by means of Lagrange's multipliers $\lambda_k, k = 1, \ldots, N$:

$$J(\mathbf{X},\mathbf{U},\mathbf{V},\lambda) = \sum_{i=1}^{c}\sum_{k=1}^{N} (\mu_{ik})^m d^2{}_{ikA} + \sum_{k=1}^{N} \lambda_k \left(\sum_{i=1}^{c} \mu_{ik} - 1 \right), \tag{A-12}$$

and by setting the gradients of J with respect to $\mathbf{U}$, $\mathbf{V}$, λ to zero (Balasko et al.,2005). If $d^2{}_{ik} > 0, \forall\, i,k$ and $m > 1$, then $(U,V) \in M_{fc} \times \mathbb{R}^{n \times c}$ will minimize equation A-12 only if

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (d_{ikA}/d_{jkA})^{2/(m-1)}}, \quad 1 \le i \le c, \quad 1 \le k \le N, \tag{A-13}$$

and

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m \mathbf{x}_k}{\sum_{k=1}^{N} \mu_{ik}^m}, \quad 1 \le i \le c. \tag{A-14}$$

This solution also satisfies the conditions A-6–A-8. Equation A-14 gives $\mathbf{v}_i$ as the weighted mean of the data items that belong to a cluster, where the weights are the membership degrees. The overall FCM algorithm is iterative through equations A-13 and A-14. Therefore, given a number of clusters c and an error tolerance of $\varepsilon > 0$, equations A-14 and A-13 are used to compute $\mathbf{v}_i$ and the partition matrix, respectively, until the partition satisfies $\|\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}\| < \varepsilon$, where $\mathbf{U}^{(l)}$ represents the partition matrix at iteration l, composed of elements μ_{ik} from equation A-13.

Partition validation indexes

By using the same notation as above, the partition index (SCI) is defined as

$$\mathrm{SCI}(c) = \sum_{i=1}^{c} \frac{\sum_{j=1}^{N} (\mu_{ij})^2 \|\mathbf{x}_j - \mathbf{v}_i\|^2}{N_{\min_{i,k}} \|\mathbf{v}_k - \mathbf{v}_i\|^2}. \tag{A-15}$$

This index is useful when comparing different partitions with the same number of clusters. A lower SCI indicates a better partition.

The Xie and Beni index (XBI) is given by

$$\mathrm{XBI}(c) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N}(\mu_{ij})^2\|\mathbf{x}_j - \mathbf{v}_i\|^2}{N_{\min_{i,j}}\|\mathbf{x}_j - \mathbf{v}_i\|^2}. \qquad \text{(A-16)}$$

This index aims at quantifying the ratio of the total variation within clusters and their separation. The optimal number of clusters should minimize the value of this index (Xie and Beni, 1991; Balasko et al., 2005).

REFERENCES

Balasko, B., J. Abonyi, and B. Feil, 2005, Fuzzy clustering and data analysis toolbox: For use with Matlab: Ph.D. thesis, University of Veszprem.

Cooper, G. R. J., 2006, Obtaining dip and susceptibility information from Euler deconvolution using the Hough transform: Computers and Geosciences, **32**, 1592–1599.

Fitzgerald, D. J., A. B. Reid, and P. McInerney, 2004, New discrimination techniques for Euler deconvolution: Computers and Geosciences, **30**, 461–469.

Florio, G., M. Fedi, and R. Pasteka, 2006, On the application of Euler deconvolution to the analytic signal: Geophysics, **71**, no. 6, L87–L93.

Galley, A. G., M. D. Hannington, and I. R. Jonasson, 2007, Volcanogenic massive sulphide deposits, *in* W. D. Goodfellow, ed., Mineral deposits of Canada: A synthesis of major deposit-types, district metallogeny, the evolution of geological provinces, and exploration methods: Geological Association of Canada, Mineral Deposits Division Special Publication No. 5, 141–161.

Geological Survey of Canada, 1996, Release of airborne geophysical survey data, Bathurst, New Brunswick: Geological Survey of Canada Open File 3294, 95 maps, scale 1:20,000 and 1:50,000.

Gerovska, D., and M. Arauzo-Bravo, 2003, Automatic interpretation of magnetic data based on Euler deconvolution with unprescribed structural index: Computers and Geosciences, **29**, 949–960.

Keating, P., and M. Pilkington, 2004, Euler deconvolution of the analytic signal and its application to magnetic interpretation: Geophysical Prospecting, **53**, 165–182.

Keating, P., M. Thomas, and F. Kiss, 2003, Significance of a high-resolution magnetic and electromagnetic survey for exploration and geologic investigations, Bathurst Mining Camp: Economic Geology Monograph 11, 783–798.

Mikhailov, V., A. Galdeano, M. Diament, A. Gvishiani, S. Agayan, S. Bogoutdinov, E. Graeva, and P. Sailhac, 2003, Application of artificial intelligence for Euler solutions clustering: Geophysics, **68**, 168–180.

Mushayandebvu, M. F., V. Lesur, A. B. Reid, and J. D. Fairhead, 2004, Grid Euler deconvolution with constraints for 2D structures: Geophysics, **69**, 489–496.

Mushayandebvu, M. F., P. van Driel, A. B. Reid, and J. D. Fairhead, 2001, Magnetic source parameters of 2D structures using extended Euler deconvolution: Geophysics, **66**, 814–823.

Nabighian, M. X., 1972, The analytical signal of 2-D magnetic bodies with polygonal cross section, its properties and use for automated anomaly interpretation: Geophysics, **37**, 507–512.

Nabighian, M., and R. O. Hansen, 2001, Unification of Euler and Werner deconvolution in three dimensions via the generalized Hilbert transform: Geophysics, **66**, 1805–1810.

Reid, A. B., J. M. Allsop, H. Granser, A. J. Millet, and I. W. Somerton, 1990, Magnetic interpretation in three dimensions using Euler deconvolution: Geophysics, **55**, 80–91.

Reid, A. B., and D. J. Fitzgerald, 2005, Hybrid Euler magnetic basement depth estimation: Bishop 3D tests: 74th Annual International Meeting, SEG, Expanded Abstracts, 671–674.

Roest, W. E., J. Verhoef, and M. Pilkington, 1992, Magnetic interpretation using 3D analytic signal: Geophysics, **57**, 116–125.

Salem, A., and D. Ravat, 2003, A combined analytic signal and Euler method (AN-EUL) for automatic interpretation of magnetic data: Geophysics, **68**, 1952–1961.

Silva, J. B. C., and V. C. F. Barbosa, 2003, 3D Euler deconvolution: Theoretical basis for automatically selecting good solutions: Geophysics, **68**, 1962–1968.

Silverman, B., 1986, Density estimation for statistics and data analysis: Chapman & Hall.

Thompson, D. T., 1982, EULDPH — A new technique for making computer-assisted depth estimates from magnetic data: Geophysics, **47**, 31–37.

Van Staal, C. R., R. A. Wilson, N. A. Rogers, L. Fyffe, J. P. Langton, S. R. McCutcheon, V. McNicoll, and C. E. Ravenurst, 2003, Geology and tectonic history of the Bathurst Supergroup, Bathurst Mining Camp, and its relationships to coeval rocks in southwestern New Brunswick and adjacent Maine — A synthesis: Economic Geology Monograph 11, 37–60.

Xie, X. L., and G. Beni, 1991, A validity measure for fuzzy clustering: IEEE Transactions on Pattern Analysis and Machine Intelligence, **13**, 841–847.