Using self-organising maps to derive lithological boundaries from geophysically-derived data in the Mt. Isa region, Queensland

Jane H Hodgkinson
CSIRO Earth Science and Resource Engineering
Pullenvale, Brisbane
Jane.Hodgkinson@csiro.au

Stephen J Fraser
CSIRO Earth Science and Resource Engineering
Pullenvale, Brisbane
Stephen.Fraser@csiro.au

Paul Donchak
Geological Survey of Queensland
Brisbane
Paul.Donchak@deedi.qld.gov.au

SUMMARY

Susceptibility and density volumes, derived from the inversion of airborne magnetic and ground-based gravity measurements for a region of the Mt Isa inlier, Queensland, Australia, were analysed using a self-organizing map (SOM) approach. Three-dimensional sub-surface voxel distributions of susceptibility and density were derived from the inversion of magnetic and gravity data using the University of British Columbia (UBC) codes. These petrophysical volumes are often difficult to interpret because of their nebulous nature, with subtle differences between adjacent volume elements. As the SOM approach uses vector quantization, it is an ideal tool to identify subtle relationships in such volumes of disparate data. The CSIRO data-mining SOM tool (SiroSOM) was used here as it was designed specifically for the analysis of such spatially-located, diverse exploration data. Our SOM analysis of the petrophysical voxels has identified (1) some structural features that are evident on the previously constructed Geological Survey of Queensland (GSQ) model of the area; (2) anomalous voxels that form coherent patterns, which may be related to mineralisation and hence be exploration targets; and (3) explicit domains that relate to lithological packages. Further work is needed to validate the SOM results; however, our analysis has shown the value of the SOM approach for analysis of such data. By using SOM, we have been able to assess petrophysical volumes to extract information related to structure and lithological packages in addition to identifying geophysical targets with potential for mineralisation.

Key words: self-organizing maps, exploration, geophysical, magnetic, gravity

INTRODUCTION

In this analysis we have used the self-organizing map (SOM) method to integrate and analyse susceptibility and density data that were derived by Mira Geoscience using University of British Columbia (UBC) algorithms. Inversion of geophysical data such as magnetic and gravity data provides petrophysical property volumes (voxels) of subsurface attributes that can be visualised in 3 dimensional (3D) models. The inverted data are typically 'nebulous' and smooth, which can cause interpretation difficulties. Typically such datasets are viewed separately or side-by-side for comparison, or they may be compared with traditional 2- or 3D models derived from seismic lines and borehole data. As an alternative to viewing and analysing petrophysical volumes in this manner, we used SOM to integrate both datasets in an unsupervised, data-driven fashion and visualise the results. SiroSOM is a spatial analysis tool developed by CSIRO for the integrated analysis of various exploration and mining datasets including multi-variate geophysical, geochemical and/or categorical data. It has therefore been used for the purposes of this study. Previously, we have used a SOM methodology to analyse petrophysical data that were derived using the UBC inversion approach (Fraser and Hodgkinson 2008; 2009). Here, SOM is used to analyse similar data from the Mt Isa area, with the aim being to enhance pre-competitive data for improved exploration options in Queensland through the Geological Survey of Queensland (GSQ).

Self-organising map methodologies were first introduced by Kohonen (1982) and are based on the principles of vector-quantization, derived in an unsupervised manner. As a non-traditional method of data analysis, the process is data-driven, allowing the integrated analysis of complex and disparate data sets such as those presented here. In a two-stage process, the data are organised to best represent the variation within and between the domains that exist. During the SOM process, the data space is first defined by the input variables. Then the data space is randomly seeded with vectors and these become 'trained' to represent the original distribution of the samples. Next, each data sample is compared with all seed-vectors within a given radius and the most similar seed-vector wins, based on a measure of vector similarity (such as cosine or Euclidean distance). The winning seed-vector and a number of its seed-vector neighbours are then modified by a given percentage so that their properties better represent that of the input sample in question. This procedure is repeated many times for each input sample and during each iteration, the radius of influence and modification percentage is reduced so the seed-vectors become more representative of the input data. Once trained the seed-vectors become 'best matching units' (BMUs), which are represented as nodes on the 'self-organised map'. The method clusters the available data into units of similarity (best matching units or BMUs) within a 'map' of nodes that provides a base for further processing, such as the Davies-Bouldin (1979) clustering method, also used here. Additionally, the method identifies data outliers described as 'Q-errors' that identify anomalies in the integrated data and are particularly useful in exploration for minerals uncommon to the surrounding country rock. More information about SOM and SiroSOM can be found for example in Kohonen, (2001, 1984, 1982) and Fraser and Dickson (2007).

SiroSOM was developed to utilise the self-organising map method of data analysis specifically for exploration and mining purposes, due to the inordinate amount of data that is produced in the industry (Fraser and Dickson, 2007) and to extract knowledge from those datasets. The method used here to integrate inverted data was previously successfully demonstrated on data sets in Canadian minerals provinces with private and publically available data. The method is tested here and compared with a previously existing Geological Survey of Queensland (GSQ) model to assess this method's value for to add value to pre-competitive data and enhance mineral exploration in Queensland. The objectives of the study were to identify whether the SOM analysis of UBC inverted, geophysically-derived data for an area in Queensland provides a model of 'coherent' groups, patterns and clusters and to identify whether SOM patterns, groups and clusters display any consistency with the pre-existing GSQ model that was built using seismic interpretation. Additionally the work was conducted to assess whether the SOM modelling could extract additional knowledge from the data. The 80 km x 100 km region was selected specifically because there are two seismic lines through the area that were used by GSQ to build their model, which can be used to text and constrain the SOM model. Additionally the area presently has a particularly high level of exploration interest.

The outputs from the SOM analysis are considered more representative of the solid geology as they result from the combined analysis of both data sets. Analysis of petrophysical properties in SOM provides an integrated, 3D view of the subsurface properties, providing a valuable alternative analysis of the datasets used. Use of this method has both confirmed pre-existing knowledge of structure and identified new features.

METHOD AND RESULTS

The original dataset consisted of 1.08 million samples with two variables: magnetic susceptibility (mag) and gravity (den). A total of 119,998 samples had null values for both variables. A set of 'xyz' files were created for both the mag and den datasets and the datasets were then combined to make one c.sv file where each point in space within the dataset had a data value for either den, mag or both. Due to the magnitude of the magnetic susceptibility and density values, logarithmic values were used to produce more definition in the analysis.

Although SOM is able to deal with datasets that have null or missing values, it does not make sense to include samples that are null in both input variables. Additionally, 495,348 samples only had the value for one of the variables. SOM is able to deal with and predict missing values, but the purpose of this study was to analyse the data where both values were present and thus integrate the data at each of those points. Therefore, all samples that had null or missing values were excluded for this feasibility study. This provided the advantage of reducing the size of the dataset to approximately 500,000 samples that allowed faster SOM processing, whilst retaining the ability to analyse a good spread of the region. Removal of these samples does not prevent SOM analysis from being valid, as SOM is able to analyse incomplete datasets. The results therefore display null regions where either or both variables were not available to process those samples, but this does not affect the resulting SOM. Processing-time was several hours ('overnight') and therefore limited the number of times that the dataset could be run. A SOM-map size of 35x27 cells using a toroid shape (allowing data to wrap around the toroid when placed on the map), was used in the SOM analysis.

During processing, samples are assigned to nodes in the SOM, based on their values being most like the node to which they have been allocated. The distance between the actual node values and the samples' values is measured and described as a quantization error (Q-error). A sample with a high Q-error is less like the node to which it was assigned, than others with a low Q-error on that same node. A high Q-error therefore represents uncharacteristic or anomalous data perhaps defining an outlier in the dataset or samples that define boundaries.

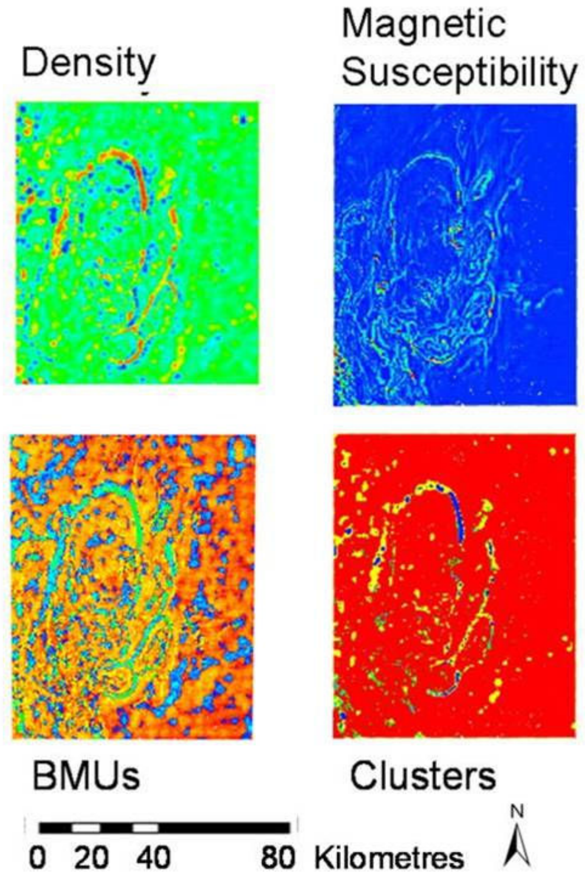

Figure 1 Comparison of visualisations available from inverted datasets and SOM processed data. Image of study area in plan-view.

A 35x27 SOM analysis produced a data-map of 1050 separate nodes, onto which samples were placed during the SOM processing iterations (each sample's BMU) and the dataset was then clustered further and the data viewed spatially. Figure 1 shows the input petrophysical property maps and corresponding SOM-derived images.. Individually, density and magnetic susceptibility show defined patterns in the model some of which coincide with one another, however, subtle variations when combined with one another are more clearly seen in the image showing the data coloured by the

SOM BMUs and further when the data has been undergone *K*-means clustering.

The resulting images showed coherency of the location of similarly-grouped and similarly-coloured samples (Figure 2a). The Davies-Bouldin *K*-means clustering method was then performed on the SOM results and each of the *K*-means clusters, also represented by a colour spatially, showed coherency of distribution (Figure 2b). This confirmed that the clusters are not randomly distributed. Finally, the most significant Q-error values also showed a non-random distribution (Figure 2c); these were grouped in the central to western region and may be of exploration interest.

When viewed individually, the datasets show different features. For example. features 1 and 2 (Figure 3a) can be seen in the magnetic data, but 2 cannot be so easily identified in the gravity data. Similarly, features 1 and 3 can be seen in the gravity data (Figure 3b) but 3 cannot be seen in the magnetic data. However, on viewing the results of the integration of the data in SOM (Figure 3c), all 3 features can be identified. Additionally, when compared with the GSQ model the SOM results confirm some structure (Figure 4) already identified by GSQ from the seismic interpretation.

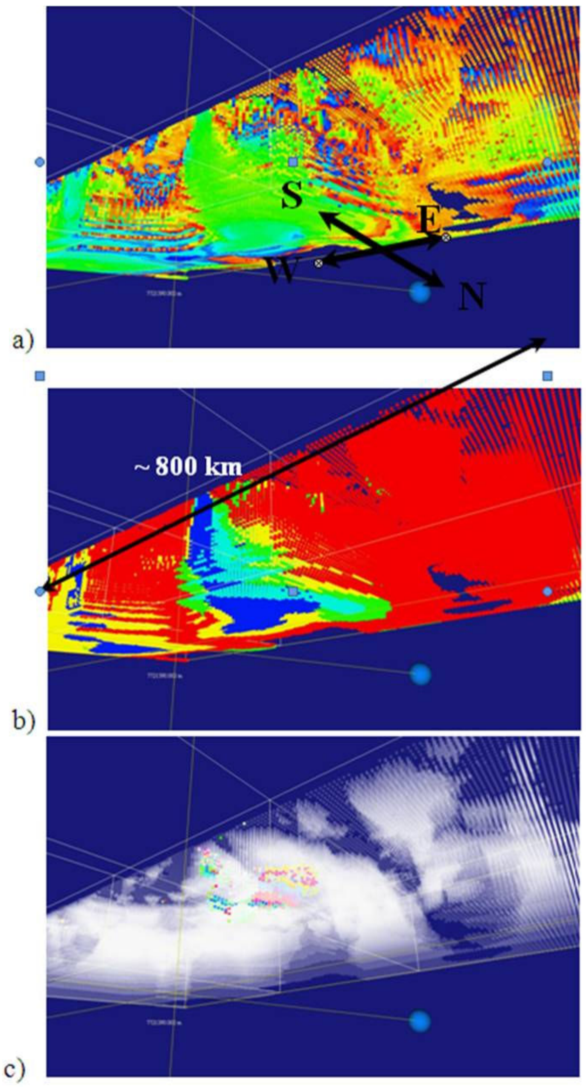

Figure 2 **BMUs or data points coloured by SOM values viewing the study area from below towards northwest. a) Nodes are coloured as assigned by SOM for each BMU showing regular patterns emerging b) nodes are coloured by assigned k-means cluster showing spatial commonalities c) nodes are coloured by value of Q-error. Only significant Q-errors are multi coloured and all others are white defining outliers and anomalies in the dataset are spatially coherent.**

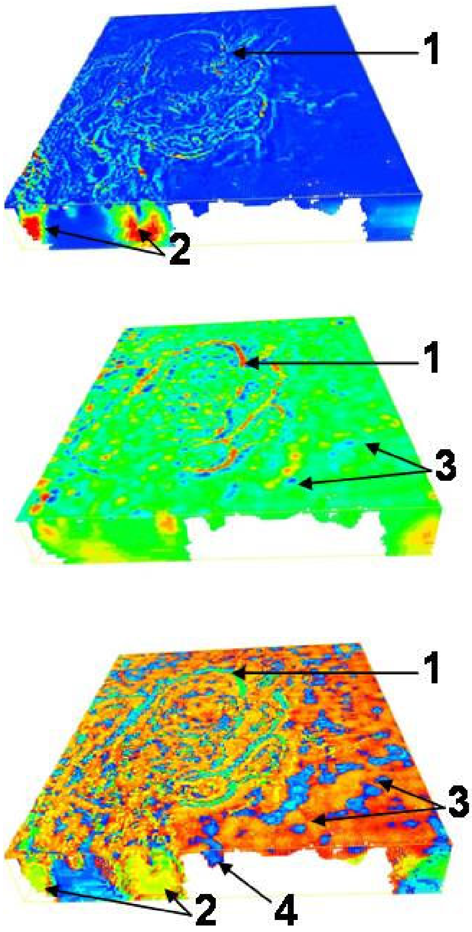

Figure 3 a) top: Image of magnetic data values in xyz space showing features 1 and 2; b) middle: image of gravity data values in xyz space showing features 1 and 3 and c) bottom: image of SOM BMUs in xyz space showing features 1, 2 and 3. All images looking north

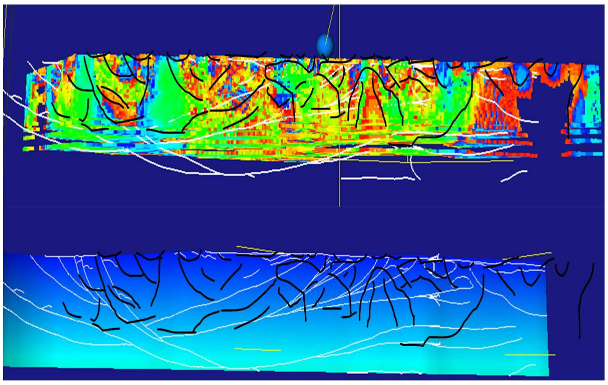

Figure 4 Comparison of the SiroSOM model trends (top) with those of the GSQ model (bottom) by overlaying GSQ structural data over the BMU-coloured model. Where some spatial trends are apparent from the clustered data.

CONCLUSIONS

SOM identified from the dataset coherent, spatial patterns relating to differences in rock type based on the petrophysical magnetic susceptibility and density data. The SOM results indicate structures through the area. The samples with high or significant Q-errors spatially close to one another identify a group of 'outliers' or anomalies that may represent samples different from the surrounding rock either due to mineralisation or other anomalous rock or mineral types atypical of rocks in the area. The location of the high Q-errors can be traced through various parts of the model by looking at slices through model; these nodes are not randomly distributed and reside at various depths indicating that the samples are 'anomalous' and may represent alternative or unusual rocks in the region in comparison with the remaining model.

The results show that SOM can be successfully used to analyse petrophysical property data; the results show coherency through the model and some features can be verified by the previous GSQ model, some new features have been described that may be of further exploration interest. The method also defines areas of the rock mass in 3 dimensions that have similarities to others, identified by SOM and k-mean clustering.

From the SOM model we can identify coherent spatial patterns in the dataset that relate to differences in lithological packages based on the petrophysical properties derived from magnetic and gravity data. The SOM results show there are structures through the area defined by the location and juxtaposition of rocks that were assigned to different nodes in the SOM map. Additionally, the non-random distribution of samples with high Q-errors suggests that there are small areas of anomalous rock character that may be different from the surrounding rock and may imply anomalous mineralisation or other uncharacteristic rock or mineral type characteristic of exploration targets. The SOM model presented here has potential for further investigation that may better define the character of the geology in this region.

REFERENCES

Davies, D.L. and Bouldin, D.W. (1979) A cluster Separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2), 224-227

Fraser, S.J., and Dickson, B.L., 2007: A New Method for Data Integration and Integrated Data Interpretation: Self-Organising Maps. In "Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration" edited by B. Milkereit, 2007, p. 907-910.

Fraser S.J., and Hodgkinson, J.H., 2009: The Analysis of Geophysically-Derived Petrophysical Volumes Using Self Organizing Maps. Invited presentation, Prospectors and Developers Association of Canada Convention 2009, March 1-5, 2009. (CSIRO Exploration & Mining Report P2009/73) : http://www.pdac.ca/pdac/conv/2009/pdf/tech-session/ts-fraser.pdf

Fraser, S.J., and Hodgkinson, J.H., 2008, An Investigation Using SiroSOM for the Analysis of QUEST Stream-Sediment and Lake-Sediment Geochemical Data, Geoscience British Columbia Canada P2009/983

Kohonen, T. (2001). Self-Organizing Maps, Third extended edition. Berlin, Heidelberg, New York, Springer

Kohonen, T. (1984). Self-Organization and Associative Memory. Berlin, Springer

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics **43**: 59-69