Application of Machine Learning Techniques to Exploration: An Example Using Self-Organising Maps for Garnet Data

Alan F. Kobussen[1], Paul D. Agnew[1], Graeme Broadbent[2]

[1] Rio Tinto Exploration, Melbourne, Australia, alan.kobussen@riotinto.com, paul.agnew@riotinto.com
[2] Rio Tinto Exploration, Perth, Australia, graeme.broadbent@riotinto.com

Introduction

The application of machine learning techniques in the earth sciences is relatively rare despite the availability of large, suitable datasets. The chemistry of diamond indicator minerals is one such dataset that has been routinely and consistently collected for the purposes of exploration and research for at least the last 50 years. There are established methodologies for interpreting these data, particularly for garnet (e.g. Gurney, 1984; Grütter et al., 2004), which are in widespread use by diamond explorers today. The existing methodologies are effective at categorising and prioritising garnet chemical data, in most cases by using binary or linear cuts in the data for one or more major-element variables. The self-organising map (SOM) is a variety of machine learning technique which organises data based on all input variables simultaneously. By applying this technique to garnet compositional dataset, relationships emerge which are not apparent in diagrams traditionally used for garnet indicator data interpretation. These relationships can be used as the basis of an exploration tool which more effectively sub-divides and prioritises garnet data from all parageneses.

The SOM Method and Dataset

The SOM is a data analysis and visualisation tool (Kohonen, 2001). It is a vector-based unsupervised clustering technique, meaning no prior knowledge of initial grouping is required. The data are organised by the SOM according to similarity using a competitive, iterative training process. The output of the SOM is a two-dimensional representation (a "map") of the multi-variate input data. The map is composed of a collection of nodes arranged in a grid. In this map, topology is preserved such that points lying close to one another on the map are more similar in composition than points lying far apart. Toroidal grids can improve results and are commonly used. Each node in the grid represents a particular n-dimensional vector, referred to as the weight vector, where n is the number of variables used to construct the map. The size of the map is determined by the user and can be varied according to task or the size of the input data. Small maps have the effect of clustering the data while large maps are more effective for visualising topology.

For this study 151,044 major-element analyses of garnet were compiled. This includes data from a wide variety of rock types in order to capture a comprehensive picture of garnet major-element compositions. Oxides of the eight major elements of the common garnet series end-members (SiO_2, TiO_2, Al_2O_3, Cr_2O_3, FeO, MnO, MgO, CaO) were used as input variables. Minor oxides, such as Na_2O, were not included due to incompleteness in the dataset and concerns about the quality and precision of some measurements. A toroidal SOM with 8064 nodes (84 by 96) was then constructed from this data using the SiroSOM software package (Fraser and Dickson, 2007).

Results

A convenient way to visualise the SOM result is by plotting the magnitudes of individual weight vector components in space. The garnet SOM has eight-dimensional weight vectors; one dimension for each of the major-elements. Overlaying these images then gives a picture of the distribution of the elements relative to one another in garnet compositional space. Maps of the Cr_2O_3 and CaO component vectors are shown as an example in Figure 1A and 1B, respectively.
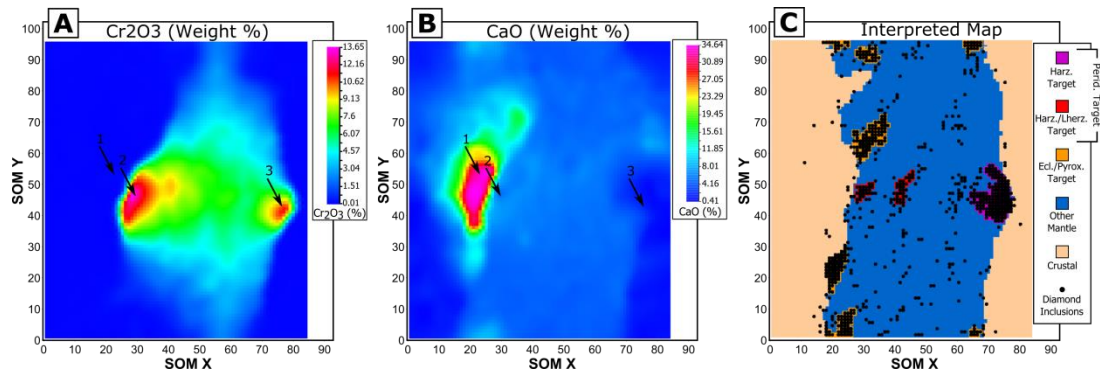
Figure 1: SOM weight vector component plots for garnet Cr_2O_3 and CaO concentrations, in weight percent (A and B), and the interpreted map (C). The garnets highest in CaO (lying in the map area indicated by arrow 1 in A and B) represent the grossular and andradite garnet end-members, nearby nodes on the map rich in Cr_2O_3 with modest CaO (indicated by arrow 2) are garnets dominated by the uvarovite molecule, and garnets poor in CaO, but rich in Cr_2O_3 (indicated by arrow 3; these garnets are also rich in MgO which is not depicted here) are dominated by the knorringite molecule. The distribution of the chemistry of garnet inclusions in diamond (C; black dots) have been used to define target compositional fields for new data (the purple, red and orange outlined areas). Due to the toroidal grid geometry, opposite edges of the map are connected (left-right and top-bottom).

The original 151,044 data points are divided across the 8064 nodes based on their Euclidian distance from the weight vector of each node. The node with the minimum distance between the weight vector and the training vector is referred to as the best matching unit (BMU). Highlighting the BMUs where specific groups of the reference data are present reveals spatial patterns on the SOM related to similarities in garnet composition. The distribution in SOM space of the compiled data for garnet inclusions in diamond is shown in Figure 1C. The diamond inclusions cluster into several coherent compositional groups, separated by zones where diamond inclusions are sparse or completely absent. These clusters have been used to define target compositional fields on the SOM map. Defining the target in this way is based on the observations that: 1) there is a definite association between these garnet compositions and diamond; and 2) coherent clustering of the majority of diamond inclusion data suggests a common process (or processes) is responsible for the formation of most diamond globally, which is reflected in the garnet composition.

To understand how effective the SOM target compositional fields are at identifying a diamondiferous source rock in comparison to existing methods, the reference data have been classified using the scheme of Grütter et al. (2004) and compared against the SOM result in Figure 2, and compiled in Table 1. In this comparison, the diamond inclusion data have been excluded to avoid bias. The proportion of G10D garnets coming from diamondiferous sources is high (84.4%), and similar to the values for the peridotitic SOM target fields (86.3%). Note the SOM target fields contain a significant lherzolitic (G9) component (Fig. 2D). The eclogitic/pyroxenitic SOM target fields define an extremely wide compositional range (Fig. 2F), but despite this the proportion of garnets from diamondiferous sources in the target field remains high (90.2%). Without reliable Na_2O data for a majority of garnet, the comparison to the scheme of Grütter et al. (2004) for eclogitic and pyroxenitic lithologies cannot be evaluated.

Group of Data	Count of Garnets from Barren Sources (Count of Sources)	Count of Garnets from Diamondiferous Sources (Count of Sources)	Count of Garnets from Sources with Unknown Diamond Content (Count of Sources)	Effectiveness (Count Diamondiferous/(Count Diamondiferous + Count Barren))*100
G10D	450 (59)	2,426 (191)	2 (1)	84.4%
Peridotitic SOM Target	323 (55)	2,033 (184)	8 (2)	86.3%
Eclogitic/Pyroxenitic SOM Target	297 (50)	2,722 (141)	428 (2)	90.2%
All Reference Data	75,407 (181)	68,066 (298)	3,649 (4)	N/A

Table 1: Effectiveness of classification technique on reference data, exclusive of diamond inclusion data.

Figure 2: A comparison of the distribution of garnet compositions on SOM and Cr_2O_3-CaO plots. A,B) Distribution of G10D garnets in SOM space and Cr_2O_3-CaO space, respectively; C,D) Distribution of peridotitic SOM target garnets in SOM space and Cr_2O_3-CaO space, respectively; E,F) Distribution of eclogitic and pyroxenitic SOM target garnets in SOM space and Ca-Mg-Fe (molar) space, respectitvely. SOM map colours same as Figure 1. The lines on the Cr_2O_3-CaO plots are the GDC of Grütter et al. (2006) and the G9/G10 dividing line of Gurney (1984).
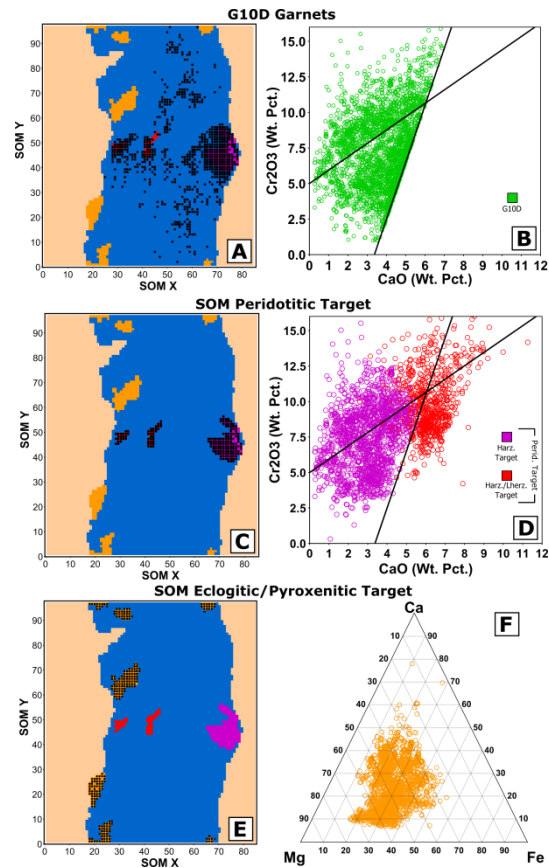
Discussion and Conclusions

Early work on garnets established the empirical link between host rock and diamond based on Cr_2O_3-CaO relationships. The SOM technique presented here is an extension of this work into eight dimensions, with the same empirical focus on diamond association. The observed clustering of diamond inclusion data has been used to define garnet target composition and results in the ability to predict whether the garnet host rock was diamondiferous with similar accuracy to existing classification schemes for peridotitic lithologies. However, the peridotitic SOM target clusters also capture a lherzolitic (G9) component while maintaining a similar accuracy to the comparison scheme here. This population of garnets may reflect the minor population of lherzolite-associated inclusions found in most diamond-bearing source rocks. It is not possible to directly compare the effectiveness of the eclogitic and pyroxenitic SOM target fields to commonly used techniques based on Na_2O, but the SOM can be shown to correctly predict a diamondiferous source with a high degree of accuracy. These factors make the garnet SOM a powerful tool for the prediction of diamond potential of a host rock without the exploration expense of fusion for microdiamonds or drill testing, particularly for indicator suites dominated by lherzolitic or eclogitic lithologies which are traditionally difficult to evaluate.

The SOM technique is based entirely on data relationships and as a result does not necessarily require the detailed petrological and mineralogical research and understanding that have developed around indicator mineral interpretation. This makes it well suited to application to other minerals with a smaller knowledge base than garnet, such as chromite and ilmenite, which also have large available datasets.

References

Fraser SJ, Dickson BL (2007) A new method for data integration and integrated data interpretation: Self-organising maps. In: Milkereit B (ed) Proceedings of Exploration 07: Fifth decennial international conference on mineral exploration. Decennial Mineral Exploration Conferences, Toronto, pp 907-910

Grütter HS, Latti D, Menzies A (2006) Cr-saturation arrays in concentrate garnet compositions from kimberlite and their use in mantle barometry. J Pet 47:801-820

Grütter HS, Gurney JJ, Menzies AH, Winter F (2004) An updated classification scheme for mantle-derived garnet, for use by diamond explorers. Lithos 77:841-857

Gurney JJ (1984) A correlation between garnets and diamonds. In: Glover JE, Harris PG (eds) Kimberlite occurrence and origin: a basis for conceptual models in exploration. Geology Department and University Extension, Publication 8, University of Western Australia, Perth, pp 143-166

Kohonen T (2001) Self-organizing maps. Springer-Verlag, Berlin