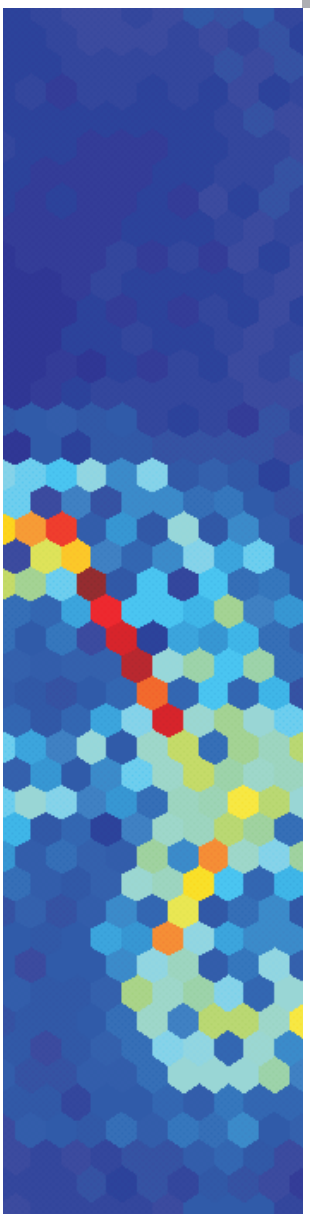
A New Method for Data Integra
Self-Organizing Maps

Stephen Fraser and Bruce Dickson

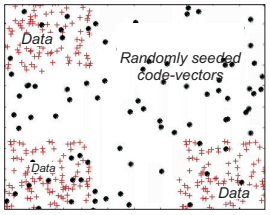


SUMMARY

The Self-Organizing Map (SOM) technique is a powerful tool for the objective analysis of complex data sets (Kohonen, 2001). SOM has been widely used for data analysis in the fields of finance, speech analysis, astronomy and more recently in petroleum well log and seismic interpretation. CSIRO has developed its own SOM toolkit tailored for use with geoscientific data, including tools for component analysis, spatial analysis and the ability to use a pre-computed SOM as a classification framework for a new dataset.

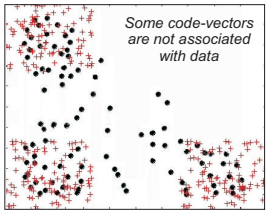
The SOM method is an unsupervized data analysis and visualization technique that is based on the principles of vector quantization. SOM makes no assumptions about statistical distributions or linear correlations of variables and therefore is suitable for analyzing the subtle and often complex relationships that are the result of geological processes. Other major strengths of the technique include the robust handling of diverse input data, categorical variables, and incomplete data.

HOW IT WORKS



Initialization. *A set of random seed vectors is initialized throughout the input data vector space.*

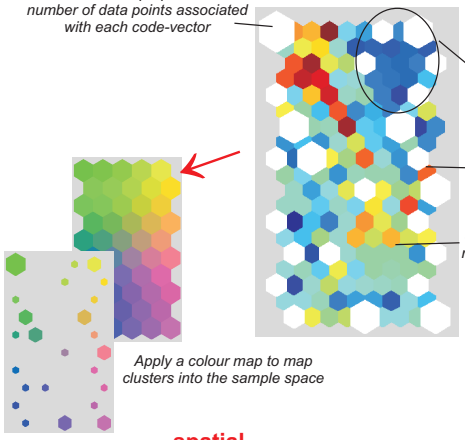
SOM training process



Self-Organization. *By a process of iterative, competitive and co-operative learning, the seed vectors "move" until they represent the structure of the input data set.*

Self-Organized mapping

White dots are proportional to number of data points associated with each code-vector



Cool colour between clusters shows they are similar

Warm colour between clusters shows they are very different

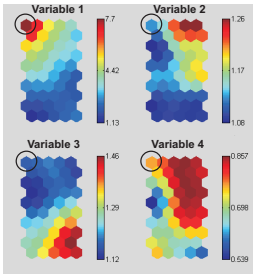
Some code-vectors are not associated with data

Mapping to the SOM. *The code-vectors are mapped onto a surface called the 'self-organized map', whereby the code-vectors are projected onto a regular grid, which is coloured according to the similarity between adjacent code-vectors. This representation highlights the relationships of nodes to each other.*

Apply a colour map to map clusters into the sample space

component analysis

Component maps to characterize each cluster

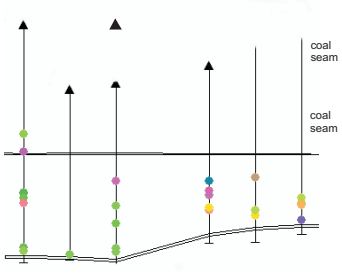


The circled node is very high in Variables 1 & 4, and low in Variables 2 & 3

Component Plot Analysis. *Because each code vector is defined in the variable space of the data set, it is possible to plot its characteristic response to each variable.*

Colour bars are stretched to represent the range of values for each variable

spatial analysis



Spatial Analysis. *A colour scheme may be applied to the map to track data clusters back into real space, or to crossplots within the dataset.*

Dickson Research Pty Ltd

For further information about SOM:

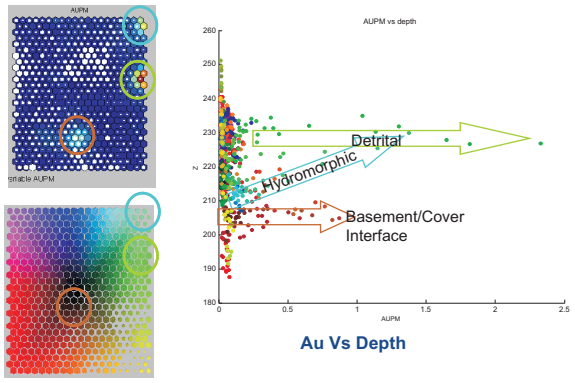
Contact Stephen Fraser
Phone +61 7 3327 4544
Email Stephen.Fraser@csiro.au
Web www.em.csiro.au

tion and Integrated Data Interpretation:

EXAMPLES

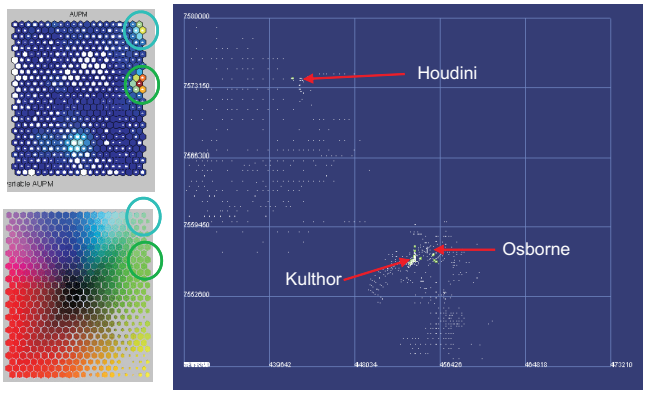
1. SOM on 40,000 geochemical samples assayed up to 13 elements. Sixty per cent of the database is 'empty.'

SOM can assist with the identification of sample populations related to "process".

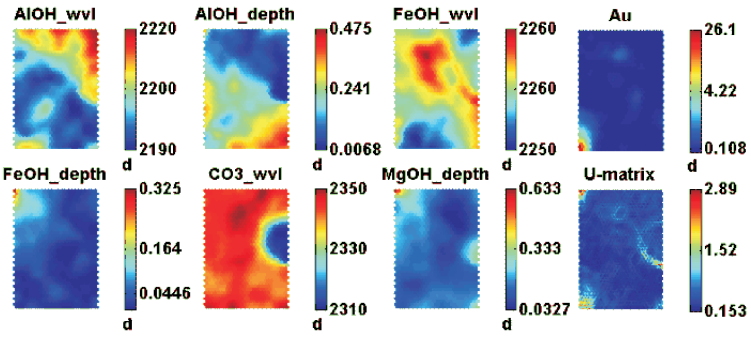


Au Vs Depth

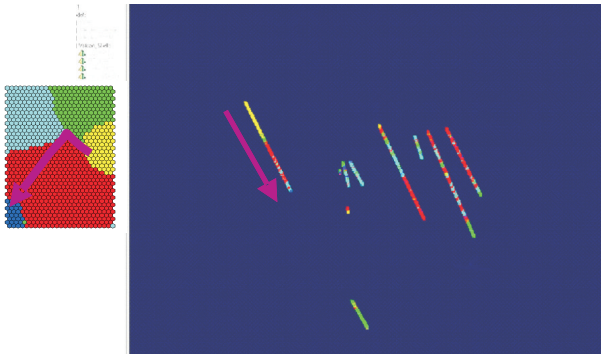
SOM can assist in target identification.



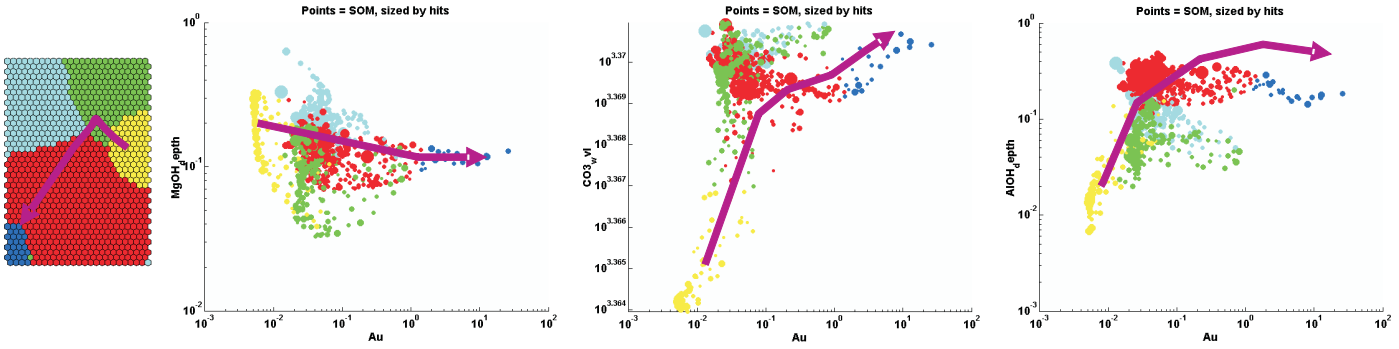
2. SOM on a Drill-hole Geochemistry and "Hylogger" spectral mineralogical database



Component Plots



SOM identifies "Vectors to Ore"



Scatter plots showing "Process of Mineralization"