Applying supervised machine learning and multiscale analysis on drill core data to improve geological logging

F. Huang[1], *E.J. Hill*[2] *and A. Otto*[3]

1. Research Scientist, CSIRO Mineral Resources, Kensington WA 6151. Email: f.huang@csiro.au
2. Principal Research Scientist, CSIRO Mineral Resources, Kensington WA 6151. Email: June.Hill@csiro.au
3. Principal Geologist, CSIRO Mineral Resources, Kensington WA 6151. Email: alex.otto@csiro.au

Keywords: machine learning, wavelet tessellation, geological logging, multivariate analysis

ABSTRACT

Geochemical data from drill core samples is used in mineral exploration to enable geologists to characterise subsurface geology. However, it is challenging to analyse these data sets rigorously and consistently. Manual data interpretation relies on the knowledge and experience of logging geologists. When multiple geologists are involved, it is possible to introduce inconsistencies in logged lithology type and scale of logging ("splitters" versus "lumpers"). Machine learning (ML) is a powerful tool for analysing high dimensional data (i.e. many variables) and large data sets (i.e. many samples). However, ML methods applied to drill hole samples do not incorporate spatial information and this can result in high misclassification rates and small-scale "noisy" results. We demonstrate the use of multiscale spatial analysis (wavelet tessellation) to incorporate spatial information into automated logging of drill core to help reduce misclassification and filter out unwanted small-scale variation in the data.

An experiment combined the use of wavelet tessellation and supervised ML to predict the rock types from two geochemistry datasets: (1) Valhalla uranium deposit in the Mount Isa Inlier, Queensland, and (2) Sunrise Dam gold mine, Western Australia. First, geochemical knowledge and statistical analysis was used to select appropriate geochemical elements for litho-geochemical classification. Second, the elements were processed using multiscale spatial analysis. Third, several ML algorithms were tested, including logistic regression, support vector machines and several tree models. We found that all ML algorithms behave worse in predicting rare rock samples than common ones. Therefore, we developed a statistical model to synthesize samples of rare rock types to overcome this issue. Results show that Random Forest has the highest accuracy of prediction. We conclude that the combination of multiscale spatial analysis (to define rock units) and Random Forests (to classify rock units) is a good method for rapid and accurate analysis of drill hole geochemistry.