A machine learning approach to geochemical mapping

Charlie Kirkwood, Mark Cave, David Beamish, Stephen Grebby, Antonio Ferreira

PII: S0375-6742(16)30098-X
DOI: doi: 10.1016/j.gexplo.2016.05.003
Reference: GEXPLO 5740

To appear in: *Journal of Geochemical Exploration*

Received date: 4 February 2016
Revised date: 16 March 2016
Accepted date: 7 May 2016

Please cite this article as: Kirkwood, Charlie, Cave, Mark, Beamish, David, Grebby, Stephen, Ferreira, Antonio, A machine learning approach to geochemical mapping, *Journal of Geochemical Exploration* (2016), doi: 10.1016/j.gexplo.2016.05.003

A machine learning approach to geochemical mapping

Charlie Kirkwood [a,*], Mark Cave [a], David Beamish [a],
Stephen Grebby [a], Antonio Ferreira [a]

[a] *British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, UK*

* Corresponding author. Tel.: +44 1159363344
Email address: cwk@bgs.ac.uk (C.W.Kirkwood)

Abstract

Geochemical maps provide invaluable evidence to guide decisions on issues of mineral exploration, agriculture, and environmental health. However, the high cost of chemical analysis means that the ground sampling density will always be limited. Traditionally, geochemical maps have been produced through the interpolation of measured element concentrations between sample sites using models based on the spatial autocorrelation of data (e.g semivariogram models for ordinary kriging). In their simplest form such models fail to consider potentially useful auxiliary information about the region and the accuracy of the maps may suffer as a result. In contrast, this study uses quantile regression forests (an elaboration of random forest) to investigate the potential of high resolution auxiliary information alone to support the generation of accurate and interpretable geochemical maps. This paper presents a summary of the performance of quantile regression forests in predicting element concentrations, loss on ignition and pH in the soils of south west England using high resolution remote sensing and geophysical survey data.

Through stratified 10-fold cross validation we find the accuracy of quantile regression forests in predicting soil geochemistry in south west England to be a general improvement over that offered by ordinary kriging. Concentrations of immobile elements whose distributions are most tightly controlled by bedrock lithology are predicted with the greatest accuracy (e.g. Al with a cross-validated R^2 of 0.79), while concentrations of more mobile elements prove harder to predict. In addition to providing a high level of prediction accuracy, models built on