PREDICTIVE GEOCHEMICAL MAPPING USING MACHINE LEARNING IN WESTERN KENYA

Olivier S. Humphrey1*, Mark Cave1, Elliott M. Hamilton1, Odipo Osano2, Diana Menya3 and Michael J. Watts1
1. Inorganic Geochemistry, Centre for Environmental Geochemistry, British Geological Survey,
Nottingham, UK
2. School of Environmental Sciences, University of Eldoret, Eldoret, Kenya
3. School of Public Health, Moi University, Eldoret, Kenya
*Corresponding author: olih@bgs.ac.uk

Abstract
Digital soil mapping is a cost-effective method for obtaining detailed information regarding the
spatial distribution of chemical elements in soils. Machine learning (ML) algorithms such as random forest (RF) models have been developed for such tasks as they are capable of modelling
non-linear relationships using a range of datasets and determining the importance of predictor
variables, offering multiple benefits to traditional techniques such as kriging.
In this study, we describe a framework for spatial prediction based on RF modelling where inverse
distance weighted (IDW) predictors are used in conjunction with auxiliary environmental covariates. The model was applied to predict the total concentration (mg kg-1) of 56 elements, soil
pH and organic matter content, as well as to assess prediction uncertainty using 466 soil samples
in western Kenya (Watts et al 2021). The results of iodine (I), selenium (Se), zinc (Zn) and soil
pH are highlighted in this work due to their contrasting biogeochemical cycles and widespread
dietary deficiencies in sub-Saharan Africa, whilst soil pH was assessed as an important parameter
to define soil chemical reactions. Algorithm performance was evaluated to determine the importance of each predictor variable and the model's response using partial dependence profiles.
The accuracy and precision of each RF model were assessed by evaluating the out-of-bag predicted
values. The IDW predictor variables had the greatest impact on assessing the distribution of soil
properties in the study area, however, the inclusion of auxiliary values did improve model performance for all soil properties.
The results presented in this paper highlight the benefits of ML algorithms which can incorporate
multiple layers of data for spatial prediction, uncertainty assessment and attributing variable importance. Additional research is now required to ensure health practitioners and the agri-community utilise the geochemical maps presented here, and the webtool, for assessing the relationship between environmental geochemistry and endemic diseases and preventable

micronutrient deficiency.

References

Watts, M., Humphrey, O., Cave, M., Osano, O., Menya, D. 2021. Western Kenya soil geochemistry.
NERC EDS National Geoscience Data Centre. (Dataset). https://doi.org/10.5285/bd1f80ef-114a-
429d-a629-20d19bacec79