

EGU2020-12378

<https://doi.org/10.5194/egusphere-egu2020-12378>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## Feasibility of using environmental covariates and machine learning to predict the spatial variability of selected heavy metals in soils

Mojtaba Zeraatpishah<sup>1</sup>, Rouhollah Mirzaei<sup>2</sup>, Younes Garosi<sup>3</sup>, Ming Xu<sup>1</sup>, Gerard B.M. Heuvelink<sup>4</sup>, Thomas Scholten<sup>5</sup>, and Ruhollah Taghizadeh-Mehrjardi<sup>5</sup>

<sup>1</sup>Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, College of Environment and Planning, Henan University, Kaifeng 475004, China (zeraatpishah@yahoo.com; mingxu@henu.edu.cn)

<sup>2</sup>Department of Environmental Sciences, University of Kashan, Kashan, Iran (rmirzaei@kashanu.ac.ir)

<sup>3</sup>Department of Soil Science, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran (unus.garosi@gmail.com)

<sup>4</sup>Wageningen University, Soil Geography and Landscape Group, 6708 11 PO Box 47, 6700 AA, Wageningen, The Netherlands (gerard.heuvelink@wur.nl)

<sup>5</sup>Soil Science and Geomorphology, Institute of Geography, Eberhard Karls University Tübingen, 72070 Tübingen, Germany (thomas.scholten@uni-tuebingen.de; ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de)

Heavy metal contamination in soil is a major environmental issue intensified by rapid industrial and population growth. Understanding the spatial distribution of soil contamination by heavy metals in the ecosystem is a necessary precondition to monitor soil health and to assess the ecological risks. The main sources of heavy metals in soil are natural and anthropogenic sources. Natural sources are typically released of heavy metals from rock by weathering and atmospheric precipitation. Anthropogenic sources are related to industrialization, rapid urbanization, agricultural practices, and military activities. We analyzed a total of 358 topsoil samples (0–30 cm) collected in Golestan province in the northeast of Iran based on a regular square grid networks with 1,700 squares each sized 2.5 km<sup>2</sup>(random sampling within the grid). From these samples, we determined the spatial distribution of Cd, Cu, Ni, Zn, and Pb using random forest (RF). A multi-spectral image (Landsat 8), and environmental derivatives calculated from terrain attributes, climatic parameters, parent material, land use maps, distances to mine sectors, main roads, industrial sites, and rivers were used as covariates to predict the spatial distribution of concentrations of heavy metals. The multi-collinearity of the predictors was examined by the variance inflation factor (VIF), and a feature selection process (genetic algorithm) was applied to avoid noise and optimize the selected input variables for the final model. The predictive accuracy of RF model was assessed by the mean prediction error (ME), root mean squared error (RMSE), and coefficient of determination ( $R^2$ ) using 5-fold cross-validation technique. The results showed that the concentration levels (mg kg<sup>-1</sup>) of Cd, Cu, Pb, Ni, and Zn varied from 0.02 to 2.75, 9.70 to 93.70, 6.80 to 114.20, 9.50 to 93.20, and 25.10 to 417.4, respectively. The best prediction performance was for Ni (RMSE=9.9 mg kg<sup>-1</sup> and  $R^2$ =56.6%), and the lowest prediction performance for Cd (RMSE=0.4 mg kg<sup>-1</sup> and  $R^2$ =28.0%). Environmental covariates that control soil moisture and water flow along with climatic factors were the most important variables to define the spatial distribution of soil heavy metals. We conclude that the RF model using easily accessible

environmental covariates is a promising, cost-effective and fast approach to monitor the spatial distribution of heavy metal contamination in soils.

**Keywords:** Heavy metals; digital soil mapping; machine learning; random forest; spatial variation; soil pollution.