Perspective
Volume 14:03, 2024

Journal of Civil and Environmental Engineering

ISSN: 2165-784X

Open Access

Spatial Distribution Prediction of Soil Heavy Metals Using Random Forest Model

Xing Xang*

Department of Environmental Sciences, Nanjing Forestry University, Nanjing 210037, China

Introduction

Soil contamination with heavy metals is a pressing environmental issue that poses significant risks to human health, agricultural productivity, and ecological balance. Heavy metals such as Lead (Pb), Cadmium (Cd), Arsenic (As), Mercury (Hg), and Chromium (Cr) are non-degradable and can persist in soils for long periods, accumulating to toxic levels. These contaminants primarily originate from industrial activities, agricultural practices, mining operations, and improper waste disposal. To manage and mitigate soil contamination, it is crucial to accurately predict the spatial distribution of heavy metals. Traditional methods of soil sampling and laboratory analysis are often time-consuming and expensive. Therefore, advanced modeling techniques, such as the Random Forest (RF) algorithm, have gained prominence for their efficiency and accuracy in predicting soil heavy metal concentrations [1].

The Random Forest model, introduced by Breiman is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or means prediction (regression) of the individual trees. This model is particularly effective for handling large datasets with complex interactions among variables. Each tree is trained on a randomly selected subset of the data (with replacement).At each split in the tree, a random subset of the features is considered for splitting, ensuring that the trees are decor related. The final prediction is made by averaging the predictions of all trees in the forest, enhancing robustness and accuracy [2,3].

Description

Important hyperparameters of the Random Forest model, such as the number of trees, maximum depth of the trees, and the number of features considered for splitting, are tuned to optimize model performance. Additionally, feature importance scores generated by the model can provide insights into the relative influence of different predictors on heavy metal concentrations. To illustrate the application of the Random Forest model, consider a case study predicting the spatial distribution of Cadmium (Cd) in agricultural soils. Soil samples are collected from various agricultural fields, and Cd concentrations are measured. Spatial data, including coordinates, land use, soil type, and topographic variables, are also gathered. The dataset is cleaned and pre-processed. Features such as soil pH, organic matter content, elevation, and distance to roads are selected. The Random Forest model is trained on the dataset. Hyper parameters are tuned to enhance model performance. The model's performance is evaluated using R^2, MSE, and RMSE metrics. Feature importance scores are analysed to identify key factors influencing Cd distribution. The trained model is used to predict Cd concentrations across the study area, generating a spatial distribution map of Cd [4].

The spatial distribution map produced by the Random Forest model highlights areas with high and low Cd concentrations. High-risk areas, such as those near industrial sites or heavily trafficked roads, can be identified for targeted remediation efforts. The feature importance analysis may reveal that soil pH and proximity to pollution sources are significant predictors of Cd concentration. The Random Forest model provides high prediction accuracy due to its ensemble nature and robustness against overfitting. Feature importance scores offer valuable insights into the factors influencing heavy metal distribution. The model can handle large datasets with numerous features, making it suitable for extensive environmental studies. The accuracy of predictions heavily depends on the quality and representativeness of the input data. Inaccurate or incomplete data can lead to erroneous predictions. Training a Random Forest model on large datasets can be computationally intensive, requiring significant processing power and memory. Soil heavy metal concentrations often exhibit spatial autocorrelation, which can violate the assumption of independent observations in the model. Techniques such as incorporating spatial lag variables or using spatially explicit models can address this issue [5].

Conclusion

The Random Forest model is a powerful tool for predicting the spatial distribution of soil heavy metals. Its ability to handle complex interactions among variables, robustness against overfitting, and provision of feature importance make it well-suited for environmental applications. Accurate spatial predictions of heavy metal concentrations can inform targeted remediation efforts, guide land use planning, and protect human health and the environment. However, the success of such predictions hinges on high-quality data, appropriate model tuning, and consideration of spatial dependencies. Future research may focus on integrating Random Forest with other machine learning techniques and geostatistical methods to further enhance prediction accuracy and address the challenges of spatial autocorrelation.

Acknowledgement

None.

Conflict of Interest

None.

References

1. Alyemeni, Mohammed Nasser and Ibrahem AA Almohisen. "Traffic and industrial activities around Riyadh cause the accumulation of heavy metals in legumes: a case study." *Saudi J Biol Sci* 21 (2014): 167-172.
2. Han, Xiran, Hao Wu, Qingyu Li and Wenrui Cai, et al. "Assessment of heavy metal accumulation and potential risks in surface sediment of estuary area: A case study of dagu river." *Mar Environ Res* (2024): 106416.

***Address for Correspondence:** Xing Xang, Department of Environmental Sciences, Nanjing Forestry University, Nanjing 210037, China; E-mail: xxang@gmail.com*

***Copyright:** © 2024 Xang X. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.*

Received: 01 May, 2024, Manuscript No. jcde-24-138696; **Editor Assigned:** 03 May, 2024, PreQC No. P-138696; **Reviewed:** 15 May, 2024, QC No. Q-138696; **Revised:** 22 May, 2024, Manuscript No. R-138696; **Published:** 29 May, 2024, DOI: 10.37421/2165-784X.2024.14.548

3. Tan, Kun, Huimin Wang, Lihan Chen and Qian Du, et al. "Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest." *J Hazard Mater* 382 (2020): 120987.

4. Tatem, Andrew J. "WorldPop, open data for spatial demography." *Sci Data* 4 (2017): 1-4.

5. Li, Zhiyuan, Zongwei Ma, Tsering Jan van der Kuijp and Zengwei Yuan, et al. "A review of soil heavy metal pollution from mines in China: Pollution and health risk assessment." *Sci Total Environ* 468 (2014): 843-853.

How to cite this article: Xang, Xing. "Spatial Distribution Prediction of Soil Heavy Metals Using Random Forest Model." J Civil Environ Eng 14 (2024): 548.