

Title: Tracing source footprints of heavy metal(oid)s in coastal soils using traditional statistical techniques and machine learning data-driven models

Author names: Abu Reza Md Towfiqul Islam , Memet Varol , Javed Mallick , Md. Abdullah-Al Mamun , Md. Yousuf Mia , Md. Abu Bakar Siddique , Md Saiful Islam , Mst. Nazneen Aktar

Abstract: Understanding the source tracing of heavy metal(oid)s (HMs) in coastal soils is paramount for efficient pollution control and safety measures. In the current study, traditional statistical techniques such as principal component analysis (PCA), principal coordinate analysis (PCoA) and machine learning data-driven models, including selforganizing maps (SOMs), conditional inference trees (CITs), ridge regression (RR) and SHapley Additive exPlanations (SHAP) were employed to detect and trace the sources of HMs in the soils along the northeast coast of Bangladesh. The concentrations of Pb (range: 11–44 mg/kg), Cd (range: 0.03–3.6 mg/kg), Mn (range: 199–981 mg/kg), and As (range: 0.13–13 mg/kg) exceeded the average shale values (ASV). This research found substantial spatial patterns of HM content and PCoA responsible for 75.46 % of total spatial variation. The SOM analysis identified natural sources for As, Ni, Fe, and Mn, and shipbreaking industry sources for Cu, Zn, and Pb in the study sites. The CIT showed that the effects of silt on soil Fe, Cd, and As, as well as the impact of soil pH on Mn and Ni, showed that these elements mostly came from geological processes. Furthermore, RR modeling provided high predictive performance for Pb, Zn, and Mn. The SHAP-based analysis quantified the importance of variables and identified potential physicochemical factors (soil pH, silt, and organic matter) that could contribute to the accumulation of HMs in soil. This paper integrates traditional and machine learning approaches to trace HM sources with a promise for future practice in other analogous areas, providing a theoretical basis for controlling soil HM contamination in coastal regions.

Keyword: Soil pollution Heavy metal(oid)s Ridge regression Self-organizing map Northeast coast of Bangladesh

[10.1016/j.marpolbul.2025.118701](https://doi.org/10.1016/j.marpolbul.2025.118701)

<https://www.sciencedirect.com/science/article/pii/S0025326X25011774>