Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.28720

Identifying candidate persistent, mobile and toxic (PMT) and very persistent and very mobile (vPvM) substances based on machine learning approaches

BIAO JIN[1], MIN HAN[1] AND HANS PETER ARP[2]

[1]Guangzhou Institute of Geochemistry, Chinese Academy of Sciences

[2]Norwegian Geotechnical Institute

Identifying persistent, mobile and toxic (PMT) substances or very persistent, very mobile (vPvM) substances among global synthetic chemicals is of primary importance for reducing contamination of drinking water. This study aims to develop a machine learning model to allow one-step prediction of candidate PMT/vPvM substances[1]. We built and compared 960 machine learning models by combining different molecular description methods, data balancing approaches and machine learning algorithms. The optimal model achieved a high accuracy of 92% for PMT/vPvM identification (i.e. positive samples) on an internal test set, and also resulted in a commendable accuracy of 90% for an external test set of chemical pollutants detected in Taihu Lake, China. Additionally, the optimal machine learning model was applied to identify the candidate PMT/vPvM substances in different sources[2]. Furthermore, we utilized Shapley additive explanations (SHAP) and causal inference to explore the relationship between PMT/vPvM properties and molecular structures. This work presents an advance of big data in-silico screening models for the identification of substances that potentially meet the PMT/vPvM criteria.

References

1. Han, M.; Jin, B.; Liang, J.; Huang, C.; Arp, H. P. H., Developing machine learning approaches to identify candidate persistent, mobile and toxic (PMT) and very persistent and very mobile (vPvM) substances based on molecular structure. *Water Research* **2023,** *244*, 120470.
2. Wang Z., Han M., **Jin B**. Identifying Candidate Persistent, Mobile, and Toxic (PMT) and Very Persistent and Very Mobile (vPvM) Substances in Shale Gas Drilling Fluids by Combining Nontarget Analysis and Machine Learning Model. Environ Sci Technol Letters 2024, 11, 2, 114–121.