Predicting Arsenic Contamination Hotspots in Abandoned River Bends in Bangladesh
A Machine Learning Approach

Biesheuvel, Julian Peter; Donselaar, Marinus Eric; Ghosh, Devanita; Lindenbergh, Roderik

Publication date
2025

Document Version
Final published version

Citation (APA)
Biesheuvel, J. P., Donselaar, M. E., Ghosh, D., & Lindenbergh, R. (2025). *Predicting Arsenic Contamination Hotspots in Abandoned River Bends in Bangladesh: A Machine Learning Approach*. Paper presented at ESA Living Planet Symposium 2025, Vienna, Austria.

Abstract

Title: Predicting Arsenic Contamination Hotspots in Abandoned River Bends in Bangladesh: A Machine Learning Approach

Authors: Julian Peter Biesheuvel, Marinus Eric Donselaar, Devanita Ghosh, Roderik Lindenbergh

Department of Geoscience and Remote Sensing, Delft University of Technology, 2628 CN Delft, The Netherlands

Session: B.04.04 Spaceborne data for the analysis of Natural Hazards and AI: new insights from Artificial Intelligence technologies and recent missions

Summary

Arsenic contamination in shallow aquifers of Holocene alluvial basins is a serious health risk affecting millions of people [1]. Detection of arsenic hotspots is a slow and tedious process based on the analysis of groundwater samples. This study improves arsenic risk prediction by incorporating geomorphological features such as oxbow lakes and clay plugs into a machine learning (ML) approach. Advances in remote sensing [2], often combined with ML, enable the efficient detection of these and other proxy features, significantly reducing reliance on labour-intensive fieldwork. By combining these features with environmental and demographic data, the approach provides more accurate and cost-effective risk assessments, enabling better-targeted interventions in vulnerable regions and supporting proactive environmental monitoring.

Introduction

Traditional geospatial interpolation methods for arsenic contamination mapping, such as Kriging and Inverse Distance Weighting, often introduce "bull's-eye" artefacts, misrepresenting true arsenic distributions. Machine learning approaches as Random Forest enhance predictions by, for example, incorporating soil types, but may still overlook crucial geomorphological features such as oxbow lakes and point bars, which act as arsenic repositories by stratigraphic trapping and groundwater diffusion processes in Holocene alluvial basins [3 – 5]. By integrating geomorphological insights with machine learning, as proposed by [6], this study enhances arsenic risk prediction, utilising a three-stage process: delineating oxbow lakes and clay plugs from satellite data, predicting arsenic concentrations based on key variables, and performing risk analysis. This approach offers more accurate risk assessments, guiding targeted mitigation strategies.

Materials and Methods

Oxbow Lake and Clay-plug Extraction

Oxbow lakes and clay plugs were identified using machine learning and algorithmic methods applied to Sentinel-1 SAR and Landsat-8 True Colour imagery via Google Earth Engine (GEE). Water bodies in the Ganges-Brahmaputra Basin were detected with Sentinel-1 VV/VH polarisation bands and the mean Normalized Difference Water Index (mNDWI) from Landsat-8. Oxbow lakes were extracted using a two-step process: rivers were excluded with Global Surface Water Explorer data, and a so-called circularity index was applied to the remaining water bodies. Clay plugs, characterised by crescent shapes and distinct perpendicular agriculture plot orientations, were detected with a You-Only-Look-Once (YOLO) deep learning model. The YOLO model is utilised for its ability to detect complex spatial patterns, such as the distinctive shapes of clay plugs.

Arsenic Concentration Prediction

Arsenic concentrations in Bangladesh were predicted using a Random Forest (RF) model, with 17,000 arsenic aquifer concentration values as ground-truth data. Predictor variables so far include precipitation, temperature, NDVI (a proxy for microbial activity in oxbow lakes and clay plugs), local terrain relief, and population density derived from satellite and climatological datasets. RF models are employed for their interpretability in understanding the relationship between features and arsenic concentrations. The model's performance was evaluated with cross-validation, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics. The predicted arsenic concentrations were validated by testing the model on samples from the original dataset that were not used during training or validation. Additionally, the model was tested on a smaller, independent dataset from a similar geomorphological environment in India that had not been seen during the training or validation phases, ensuring the model's robustness and generalizability.

Arsenic Risk Prediction

Arsenic risk was classified based on predicted concentrations, relative terrain height, and local population density. High population density and predicted high arsenic concentrations indicate immediate risk. In contrast, low population density and high arsenic concentrations signal potential future risks, especially due to climate change or increased organic matter. Risk levels were assigned with rule-based or unsupervised machine learning clustering methods.

Initial and Expected Results

Initial results focus on successfully applying YOLO to detect and delineate geomorphological features such as clay plugs from satellite imagery. Early tests have demonstrated the model’s ability to identify crescent-shaped landforms with varying degrees of accuracy depending on the quality of the spatial and temporal resolution of the input data. Additionally, oxbow lakes have been successfully extracted using SAR data and the mNWDI method.

The next steps for the YOLO method involve expanding the approach to include false colour imagery, which may provide more contrast and better highlight the crescent-shaped features associated with clay plugs. Additionally, the model will be trained with a larger and more diverse dataset, including a dedicated validation and test set, to ensure robust evaluation and avoid overfitting. These improvements are expected to enhance the accuracy and generalizability of the YOLO-based detection of geomorphological features in satellite imagery.

Conclusions

This research introduces a more precise approach for identifying arsenic hotspots using a limited set of proxy features, eliminating the need for costly and time-intensive fieldwork. Once trained, the model can be fine-tuned and applied to other regions with comparable geomorphological characteristics, making it adaptable and scalable for global applications. Integrating geomorphological features such as oxbow lakes and point bars with machine learning models improves arsenic risk prediction accuracy. This approach enables fast and accurate arsenic monitoring and supports targeted mitigation strategies safeguarding public health, particularly in vulnerable regions. Future work will focus on refining model parameters, expanding datasets, and incorporating additional environmental factors to enhance the model's robustness and applicability.

Impact

This research presents a novel, cost-effective approach to mapping arsenic contamination by leveraging machine learning and geomorphological features. Utilising proxy data reduces the need for extensive fieldwork, making arsenic risk assessment more accessible, especially in resource-limited regions. The integration of geomorphological insights enhances the accuracy of arsenic predictions, which can lead to better-targeted interventions and policy decisions, particularly in areas vulnerable to arsenic contamination, such as the Ganges-Brahmaputra Basin. This work contributes to the growing field of environmental monitoring, providing proactive risk management in the context of water quality and public health.

Acknowledgements

The authors thank the Environment and Population Research Center (EPRC), Dhaka, Bangladesh, for providing the dataset containing arsenic concentrations in Bangladesh.

References

[1] Rahaman, M. S., Mise, N., & Ichihara, S. (2022). Arsenic contamination in the food chain in Bangladesh: A review on health hazards, socioeconomic impacts, and implications. *Hygiene and Environmental Health Advances, 2*, 100004. https://doi.org/10.1016/j.heha.2022.100004

[2] Agarwal, V., Kumar, M., Panday, D. P., Zang, J., & Munoz-Arriola, F. (2024). Unlocking the potential of remote sensing for arsenic contamination detection and management: Challenges and perspectives. *Current Opinion in Environmental Science & Health, 100578*. https://doi.org/10.1016/j.coesh.2024.100578

[3] Kumar, S., Ghosh, D., Donselaar, M., Burgers, F., & Ghosh, A. (2021). Clay-plug sediment as the locus of arsenic pollution in Holocene alluvial-plain aquifers. Catena, 202, 105255. https://doi.org/10.1016/j.catena.2021.105255

[4] Donselaar, M. E., Bhatt, A. G., & Ghosh, A. K. (2016). On the relation between fluvial-deltaic flood basin geomorphology and the widespread occurrence of arsenic pollution in shallow aquifers. The Science of the Total Environment, 574, 901–913. https://doi.org/10.1016/j.scitotenv.2016.09.074

[5] Ghosh, D., Kumar, S., Donselaar, M. E., Corroto, C., & Ghosh, A. K. (2020). Organic carbon transport model of abandoned river channels: A motif for floodplain geomorphology influencing biogeochemical swaying of arsenic. *The Science of the Total Environment, 762*, 144400. https://doi.org/10.1016/j.scitotenv.2020.144400

[6] Donselaar, M. E., Khanam, S., Ghosh, A. K., Corroto, C., & Ghosh, D. (2024). Machine-learning approach for identifying arsenic-contamination hot spots: The search for the needle in the haystack. *ACS ES&T Water, 4*(8), 3110–3114. https://doi.org/10.1021/acsestwater.4c00422