Finding the needle in the haystack
Machine learning approach in the search for arsenic hotspots

Donselaar, M.E.; de Aguiar Paniago de Sousa, L.; Kumar, S.; Khanam, S.; Ghosh, A.K.; Corroto, C.; Ghosh, D.

Publication date
2024

Document Version
Final published version

Citation (APA)
Donselaar, M. E., de Aguiar Paniago de Sousa, L., Kumar, S., Khanam, S., Ghosh, A. K., Corroto, C., & Ghosh, D. (2024). *Finding the needle in the haystack: Machine learning approach in the search for arsenic hotspots*. Paper presented at 9th International Congress & Exhibition on Arsenic in the Environment 2024, Bhubaneshwar, India.

Finding the needle in the haystack: Machine learning approach in the search for arsenic hotspots

M.E. Donselaar[1,2], L. de Aguiar Paniago de Sousa[2], S. Kumar[1], S. Khanam[3], A.K. Ghosh[4], C. Corroto[5], D. Ghosh[6]

[1]Department of Applied Geoscience and Engineering, Delft Univ. of Technology, Delft, The Netherlands
[2]Dept. of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium
[3]Environment and Population Research Center (EPRC), Mohakhali, Dhaka, Bangladesh
[4]Mahavir Cancer Sansthan and Research Centre, Patna (Bihar), India
[5]Universidad de Buenos Aires (UBA), Argentina
[6]Water Management Department, Delft Univ. of Technology, Delft, The Netherlands

ABSTRACT: Groundwater contamination with naturally-occurring arsenic (As) poses a serious health threat of global proportions. Implementation of focused and sustainable As-mitigation measures is hampered by the inability to pinpoint enigmatic As-hotspots in the vast area of continental alluvial basins with elevated toxicity risk. The catalyser role of invasive vegetation in anoxic fluvial oxbow lakes in combination with microbial metabolism processes are key in mobilizing As from its solid state into the shallow aquifer domain and accumulation in porous and permeable fluvial point-bar sediment. This insight opens up the opportunity for a cross-disciplinary approach to construct predictive machine learning-based object-based geospatial models to locate As-hotspots by the analyses of (1) satellite imagery of alluvial geomorphology, (2) oxbow-lake vegetation density, and (3) ground-truth databases of oxbow-lake/aquifer biogeochemistry and fluvial sedimentology.

1 INTRODUCTION

Aquifer contamination with naturally-occurring (geogenic) arsenic (As) poses a serious health threat of global proportions. Research into the occurrence and location of toxic natural As levels in drinking and irrigation water, and in the food chain, provided insight into its provenance, accumulation in sedimentary basins, and the health risks of As ingestion. A bloom of potential mitigation measures have been proposed, ranging from in-situ chemical reactive processes to immobilize arsenic, filtration methods, and social awareness programs for the affected population.

The enigmatic nature of As-hotspot occurrence is the large remaining challenge for a successful application of sustainable mitigation measures. In terms of As-prone surface area and number of people in potential harm, Holocene alluvial plains are by far the largest hazard-prone environment. To date, attempts to map As-occurrence focused on geostatistical interpolation of discrete As-concentration measurements from tube wells, thereby disregarding the critical role of alluvial geomorphology, lithofacies heterogeneity and inherent porosity-permeability anisotropy in the efficiency of groundwater recharge, and in As accumulation in stratigraphic traps (Fig. 1).

The advent of machine learning techniques opens up the opportunity to construct predictive GIS-based geospatial models to localize As hotspots in the large extension of Holocene alluvial plains. For this, it is crucial to define a set of correct predictor variables that fully take into account the geomorphological and geological anisotropy. In this paper a cross-disciplinary approach is presented based on field studies of As-hotspot locations in Bihar and West Bengal (India) and Narail District (Bangladesh), which defines the choice of predictor variables.

2 METHODS AND DATA

The spatial distribution of lithofacies and porosity-permeability anisotropy was modelled by correlating fully-cored boreholes in oxbow lakes and adjacent point bars. Satellite imagery analysis yielded the alluvial geomorphology and seasonal evolution of oxbow-lake invasive vegetation cover. Biogeochemical analyses of sediment cores, groundwater and lake water gave insight in the catalyser role of microbial metabolism processes in anoxic oxbow lakes in mobilizing arsenic from its solid state and subsequent accumulation in permeable alluvial-plain aquifers (Ghosh et al., 2021). Ground-truth As-concentration databases served to corroborate the occurrence, migration paths and accumulation.

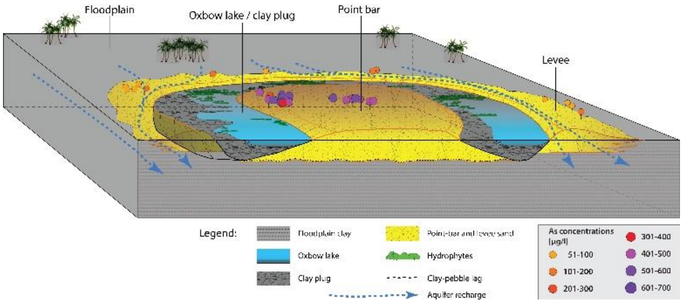

Figure 1 - Conceptual geospatial model for the relation of regional groundwater flow and As entrapment, in which permeable fluvial point-bar sand serves as a four-way geomorphological As trap by-passed by aquifer recharge flow. From: Ghosh and Donselaar (2023).

3 RESULTS AND DISCUSSION

3.1 *Geomorphological model*

Borehole correlation showed that point-bars are (1) sand-prone, porous and permeable geomorphological units with (2) inclined internal permeability baffles in its upper part, and (3) enclosed by oxbow lakes filled with fine-grained sediment in successive floods and converted to impermeable clay plugs (Fig. 1).

3.2 *Oxbow-lake vegetation: source of biomass*

Eichhornia crassipes sp. (water hyacinth) and *Hydrilla verticillata sp.* (water thyme) are the predominant macrophyte species in (sub)tropical, and in warm to cold water oligotrophic and eutrophic oxbow lakes, respectively. The macrophytes have high growth rates as they assimilate nutrients from sediment and the water column at variable rates according to seasonal rainfall and surface runoff fluctuation. Carboxyl and hydroxyl groups in the cell walls of plant leaves and stems adsorb metal cations in the solution, and both macrophytes are known As accumulators. Ghosh et al. (2021) documented macrophyte coverage of 89% of a single oxbow-lake surface of 4.37 km^2 between two successive monsoon periods, when plants are almost entirely eradicated by the flood waters and added as large volume of biomass in the lake bottom sediment. High concentrations of plants, nutrients and metal ions are a source of decomposable labile organic matter in an oxygen-depleted lake bottom that favours microbial digestion and utilization, and arsenic mobilization.

3.3 *Microbial-facilitated reductive dissolution*

Microbial analyses reveal that oxygen-deprived oxbow lake and clay-plug deposits are rich in easily decomposable organic material, to serve as highly-effective As-release loci through microorganism-facilitated reductive dissolution. Oxbow lakes are natural organic-carbon repositories, which gathers as High Molecular Weight Dissolved Organic Carbon (HMW-DOC) and as human-derived organic matter. The high HMW:TOC ratio and low overall TOC concentration, pointing to microbial breakdown and consumption of LMW. Faeces-derived markers like *Coprostanol* and *Coprostenol* indicate anthropogenic enrichment of the subsurface water, creating ideal conditions for methane-producing microbe growth that can operate with or without oxygen. As this pool of Dissolved Organic Matter (DOM) travels into aquifers, a separation based on molecular weight and resistance to breakdown occurs (Fig. 2). HMW-DOM buildup may lead to larger oxidation of As(V) and release in oxygen-depleted aquifer systems. Dissolved As(III) then moves to adjacent sandy porous and permeable point bars by diffusion and advection along the porosity-permeability gradient, driven by gravity and clay compaction.

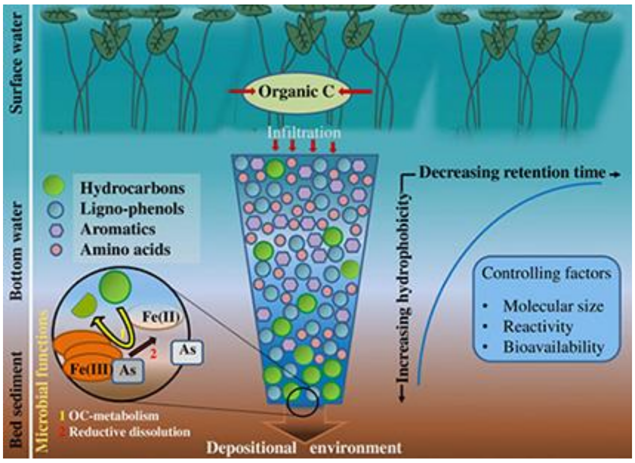

Figure 2 – Oxbow-lake chromatographic model shows the combined roles of vegetation and microbial metabolism in the release of As. From: Ghosh et al. (2021).

3.4 *Discussion*

As-hotspots are identified from the interaction of oxbow-lake vegetation-derived biomass and anoxic microbial metabolism efficiency in anoxic oxbow-lake enclosed point-bar units. The point-bar/oxbow-lake geomorphological element serves as blueprint for predictive machine-learning geospatial object-based As-hotspot models, with GIS-based satellite image data analysis of alluvial geomorphology, vegetation density analysis, climate, and population density as key predictor variables, and geological ground-truth As-concentration databases, biogeochemical and sedimentological information as training sets.

4 CONCLUSIONS

Sand-prone fluvial point bars and vegetation-rich, clay-filled oxbow lakes form a linked geomorphological element, identified as a source-to-sink ecosystem for As release and entrapment by the interaction of high biomass production and microbial reductive dissolution in the anoxic oxbow-lake setting. The identification of As-hotspot locations forms the basis for a machine-learning approach that aims at basin-wide object-based hotspot mapping by integrated analysis of GIS-based alluvial geomorphology, oxbow-lake vegetation density, and population as key predictor variables supported by ground-truth datasets.

REFERENCES

Ghosh, D., Donselaar, M.E., 2023. Predictive geospatial model for arsenic accumulation in Holocene aquifers based on interactions of oxbow-lake biogeochemistry and alluvial geomorphology. Science of the Total Environment 856, 158952.

Ghosh, D., Kumar, S., Donselaar, M.E., Corroto, C., Ghosh, A.K., 2021. Organic Carbon transport model of abandoned river channels - A motif for floodplain geomorphology influencing biogeochemical swaying of arsenic. Science of the Total Environment 762, 144400.