

Title: Using unsupervised machine learning models to drive groundwater chemistry and associated health risks in Indo-Bangla Sundarban region

Author Names: Jannatun Nahar Jannat , Abu Reza Md Towfiqul Islam , Md Yousuf Mia , Subodh Chandra Pal , Tanmoy Biswas , Most Mastura Munia Farjana Jion , Md Saiful Islam , Md Abu Bakar Siddique , Abubakr M. Idris , Rahat Khan , Aznarul Islam , Tapos Kormoker , Venkatramanan Senapathi

Abstract: Groundwater is an essential resource in the Sundarban regions of India and Bangladesh, but its quality is deteriorating due to anthropogenic impacts. However, the integrated factors affecting [groundwater chemistry](#), source distribution, and health risk are poorly understood along the Indo-Bangla coastal border. The goal of this study is to assess [groundwater chemistry](#), associated driving factors, source contributions, and potential non-carcinogenic health risks (PN-CHR) using unsupervised machine learning models such as a self-organizing map (SOM), [positive matrix factorization](#) (PMF), ion ratios, and Monte Carlo simulation. For the Sundarban part of Bangladesh, the SOM clustering approach yielded six clusters, while it yielded five for the Indian Sundarbans. The SOM results showed high correlations among Ca^{2+} , Mg^{2+} , and K^{+} , indicating a common origin. In the Bangladesh Sundarbans, mixed water predominated in all clusters except for cluster 3, whereas in the Indian Sundarbans, $\text{Cl}^{-}\text{-Na}^{+}$ and mixed water dominated in clusters 1 and 2, and both water types dominated the remaining clusters. Coupling of SOM, PMF, and ionic ratios identified rock weathering as a driving factor for groundwater chemistry. Clusters 1 and 3 were found to be influenced by mineral dissolution and geogenic inputs (overall contribution of 47.7%), while agricultural and [industrial effluents](#) dominated clusters 4 and 5 (contribution of 52.7%) in the Bangladesh Sundarbans. Industrial effluents and agricultural activities were associated with clusters 3, 4, and 5 (contributions of 29.5% and 25.4%, respectively) and [geogenic sources](#) (contributions of 23 and 22.1% in clusters 1 and 2) in Indian Sundarbans. The probabilistic health risk assessment showed that NO_3^{-} poses a higher PN-CHR risk to human health than F^{-} and As, and that potential risk to children is more evident in the Bangladesh Sundarban area than in the Indian Sundarbans. Local authorities must take urgent action to control NO_3^{-} emissions in the Indo-Bangla Sundarbans region.

Keywords: Groundwater chemistry Sundarbans Indo-Bangla coastal zone Self-organizing map (SOM) Positive matrix factorization (PMF) Unsupervised machine learning

<https://www.sciencedirect.com/science/article/abs/pii/S0045653524001103>

<https://doi.org/10.1016/j.chemosphere.2024.141217>