

Title: Predicting groundwater phosphate levels in coastal multi-aquifers: A geostatistical and data-driven approach

Author: Md. Abdullah-Al Mamun, Abu Reza Md Towfiqul Islam, Mst. Nazneen Aktar, Md Nashir Uddin, Md. Saiful Islam, Subodh Chandra Pal, Aznarul Islam, A.B.M. Mainul Bari, Abubakr M. Idris, Venkatramanan Senapathi

Abstract: The groundwater (GW) resource plays a central role in securing water supply in the coastal region of Bangladesh and therefore the future sustainability of this valuable resource is crucial for the area. However, there is limited research on the driving factors and prediction of phosphate concentration in groundwater. In this work, geostatistical modeling, self-organizing maps (SOM) and data-driven algorithms were combined to determine the driving factors and predict GW phosphate content in coastal multi-aquifers in southern Bangladesh. The SOM analysis identified three distinct spatial patterns: K^+ +single bond Na^+ +single bond pH , Ca^{2+} +single bond Mg^{2+} +single bond NO_3^- , and HCO_3^- +single bond SO_4^{2-} +single bond PO_4^{3-} +single bond F^- . Four data-driven algorithms, including CatBoost, Gradient Boosting Machine (GBM), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR) were used to predict phosphate concentration in GW using 380 samples and 15 prediction parameters. Forecasting accuracy was evaluated using RMSE, R^2 , RAE, CC, and MAE. Phosphate dissolution and saltwater intrusion, along with phosphorus fertilizers, increase PO_4^{3-} content in GW. Using input parameters selected by multicollinearity and SOM, the CatBoost model showed exceptional performance in both training (RMSE = 0.002, MAE = 0.001, R^2 = 0.999, RAE = 0.057, CC = 1.00) and testing (RMSE = 0.001, MAE = 0.002, R^2 = 0.989, RAE = 0.057, CC = 0.998). Na^+ , K^+ , and Mg^{2+} significantly influenced prediction accuracy. The uncertainty study revealed a low standard error for the CatBoost model, indicating robustness and consistency. Semi-variogram models confirmed that the most influential attributes showed weak dependence, suggesting that agricultural runoff increases the heterogeneity of PO_4^{3-} distribution in GW. These findings are crucial for developing conservation and strategic plans for sustainable utilization of coastal GW resources.

Keywords: Groundwater (GW), Phosphate Concentration, Coastal Regions, Geostatistical Modeling, Self-Organizing Maps (SOM), Data-Driven Algorithms

DOI: <https://doi.org/10.1016/j.scitotenv.2024.176024>

<https://www.sciencedirect.com/science/article/abs/pii/S0048969724061801>