Predicting Base Flow Nitrate Concentrations in Streams Using Machine Learning

ANTHONY J. TESORIERO[1*], SUSAN A. WHERRY[2] AND SILVIA TERZIOTTI [3]

[1]U.S. Geological Survey, Portland, OR, 97201, USA, tesorier@usgs.gov
[2]U.S. Geological Survey, Portland, OR, 97201, USA, swherry@usgs.gov
[3]U.S. Geological Survey, Raleigh, NC, 27607, USA, seterzio@usgs.gov

New Approach for Predicting Water Quality

Application of machine-learning techniques to existing monitoring data has recently been used to broadly assess aquifer vulnerability at regional and national scales [1,2]. Model results can also be used to better understand the environmental processes and land use practices responsible for observed conditions. In this study, machine learning is used to predict base flow nitrate concentrations in streams in the Chesapeake Bay Watershed in the United States. Elevated nitrate concentrations in streams and rivers in the Chesapeake Bay watershed have adversely affected the ecosystem health of the bay. Much of this nitrate is derived from groundwater that discharges to streams as base flow.

Model Results

Boosted regression trees (BRT) were used to relate measured nitrate concentrations in base flow (n=156) to explanatory variables describing nitrate sources, geology, and soil and catchment characteristics. From these relations, a BRT model was developed to predict base flow nitrate concentrations in streams throughout the Chesapeake Bay watershed. Riparian canopy, organic carbon content in soils and redox conditions in groundwater were among the most important variables in the model, emphasizing the importance of natural attenuation, such as denitrification and uptake by plants, on base flow nitrate concentrations. The BRT model was more accurate than a similarly constructed multiple linear regression model.

[1] Tesoriero et al., (2017), *Water Resources Research* **53**, https://doi.org/10.1002/2016WR020197.
[2] Koch et al. (2019), *Water Resources Research* **55,** https://doi.org/10.1029/2018WR023939.