

A machine learning approach to water quality forecasts and sensor network expansion: Case study in the Wabash River Basin, USA

Tyler Balson¹ and Adam Ward²

¹Indiana University Bloomington

²Indiana University

September 25, 2021

Abstract

Midwestern cities require forecasts of surface nitrate loads to bring additional treatment processes online or activate alternative water supplies. Concurrently, networks of nitrate monitoring stations are being deployed in river basins, co-locating water quality observations with established stream gauges. Here, we construct a synthetic data set of stream discharge and nitrate for the Wabash River Basin - one of the U.S.'s most nutrient polluted basins - using the established Agro-IBIS model. While real-world observations are limited in space and time, particularly for nitrate, the synthetic data set allows for sufficiently long periods to train machine learning models and assess their performance. Using the synthetic data, we established baseline 1-day forecasts for surface water nitrate at 12 cities in the basin using support vector machine regression (SVMR; RMSE 0.48-3.3 mg/L). Next, we used the SVMRs to evaluate the improvement in forecast performance associated with deployment of additional sensors. Synthetic data enable us to quantitatively assess the expected value of an additional nitrate sensor being deployed, which is, of course, not possible if we are limited to the present observational network. We identified the optimal sensor placement to improve forecasts at each city, and the relative value of sensors at all possible locations. Finally, we assessed the co-benefit realized by other cities when a sensor is deployed to optimize a forecast at one city, finding significant positive externalities in all cases. Ultimately, our study explores the potential for AI to make short-term predictions and provide an unbiased assessment of the marginal benefit and co-benefits to an expanded sensor network. While we use water quantity in the Wabash River Basin as a case study, this approach could be readily applied to any problem where the future value of sensors and network design are being evaluated.

Hosted file

balsonwardML-maintext.docx available at <https://authorea.com/users/436838/articles/538886-a-machine-learning-approach-to-water-quality-forecasts-and-sensor-network-expansion-case-study-in-the-wabash-river-basin-usa>

figures/fig1/fig1-eps-converted-to.pdf

figures/Fig2/Fig2-eps-converted-to.pdf

figures/fig3/fig3-eps-converted-to.pdf

figures/Fig4/Fig4-eps-converted-to.pdf

Hosted file

Table-1.xlsx available at <https://authorea.com/users/436838/articles/538886-a-machine-learning-approach-to-water-quality-forecasts-and-sensor-network-expansion-case-study-in-the-wabash-river-basin-usa>

Hosted file

Table-2.xlsx available at <https://authorea.com/users/436838/articles/538886-a-machine-learning-approach-to-water-quality-forecasts-and-sensor-network-expansion-case-study-in-the-wabash-river-basin-usa>