National modeling of lake nutrient & chlorophyll-*a* concentrations **to characterize HAB risk**

Meredith Brehob, Robert Sabo, & Michael Pennino

Harmful algal blooms (HABs) occur when excessive algal growth leads to negative outcomes such as unsightly and odorous waters, oxygen depletion, and toxin production. Through these mechanisms, HABs detract from the freshwater ecological services that humans rely on (e.g., recreation, fisheries, and safe drinking water). It is broadly understood that nutrient pollution leads to eutrophication of waterbodies and growth of HABs. However, because HABs occur sporadically and are difficult to monitor, it is often not possible to match observed HAB events with corresponding nutrient measurements. This mismatch makes it difficult to identify waterbodies susceptible to HABs and the anthropogenic and environmental drivers responsible.

To advance HABs research efforts, we leveraged large environmental datasets to predict nutrient conditions and chlorophyll-*a* concentrations (an indicator of HABs) in lakes that do not have robust water quality datasets. Total nitrogen, total phosphorus, and chlorophyll-a data from the 2007 and 2012 NLA surveys were paired with watershed-scale weather (PRISM Group), terrestrial net primary productivity (NASA Earth Observations), annual nutrient input (e.g., farm fertilizer and atmospheric deposition; EPA's National Nutrient Inventory), and landscape (e.g., soil properties and land cover; LakeCat Data) datasets. We then used this data in machine learning models to investigate watershed drivers of nutrient and chlorophyll-*a* concentrations and to characterize the risk of HABs in lakes across the United States (US) (Brehob et al. 2024).

To broaden and enhance the accuracy of chlorophyll-*a* predictions, we developed a modelling structure which used predicted lake nitrogen and phosphorus concentrations as indicators of chlorophyll-*a*. This approach was applied to a national dataset of watershed nutrient input data, weather parameters, and landscape factors to make nutrient and chlorophyll-*a* predictions for 112,023 lakes. In addition to watershed-specific lake predictions, we estimated lake nutrient and chlorophyll-*a* concentrations across the entire landscape. These broader landscape predictions are powerful because they provide insights into eutrophication trends across the U.S., including the smaller headwater lakes that typically lack watershed data coverage.

The models demonstrated strong predictive power, explaining 65 percent of the variation in total nitrogen, 62 percent in total phosphorus, and 68 percent in chlorophyll-*a* across the 112,023 lakes. As expected, based on our common understanding of lake eutrophication dynamics, lake chlorophyll-*a* concentrations were primarily driven by total nitrogen and total phosphorus concentrations. Interestingly, lake depth also plays an important role. As lake depth increased, nutrient and chlorophyll-a concentrations tended to decrease. While there are many ways lake depth influences nutrient and chlorophyll-*a* concentrations, the most important is likely tied to a longer water residence time. Deeper lakes tend to hold water longer, allowing more time for natural nutrient cycling, sedimentation, and processing to occur, leading to lower nutrient levels and, subsequently, lower chlorophyll-*a* concentrations.

After accounting for lake depth in the nutrient models, agricultural inputs largely drove lake nutrient concentrations, with higher predicted nutrient concentrations for lakes in major US agricultural regions (bottom panel of Figure 1). Generally, agricultural variables such as fertilizer inputs and soil agricultural erodibility factors tend to increase lake nutrient and chlorophyll-*a* concentrations, while variables such as lake depth and watershed runoff attenuate nutrient loads and lower lake chlorophyll-*a* concentrations.

This predictive modeling work leverages some of the most advanced publicly available estimates of weather conditions, terrestrial net primary production, anthropogenic nutrient inputs, soil factors, and in-lake characteristics to predict nutrient and chlorophyll-*a* concentrations in more than 100,000 lakes across the contiguous US. These results (1) provide insight into the drivers of eutrophication and HABs in lakes at a national scale and (2) identify lakes and headwater areas that are at risk for HABs. The predicted nutrient and chlorophyll-*a* estimates are publicly available and may be used by managers tasked with protecting American waterways to identify areas of concern from HABs and assist in addressing nutrient pollution challenges.

In addition to its management relevance, this work represents a broader, rapidly growing trend in environmental science: harnessing big data and machine learning to predict, forecast, and interpret major challenges at continental scales. These efforts rely on high-quality national datasets like the US EPA's National Aquatic Resource Surveys, providing dependable data from probabilistic surveys of the nation's coasts, lakes, rivers, and wetlands; the National Nutrient Inventory, delivering detailed estimates of anthropogenic nitrogen and phosphorus inputs and outputs; and the StreamCat and LakeCat datasets, providing reach- and watershed-scale landscape metrics for every lake and stream in the contiguous US. As HABs continue to compromise our water resources, scalable tools built on these foundational EPA resources and other national datasets are essential. This study and the associated

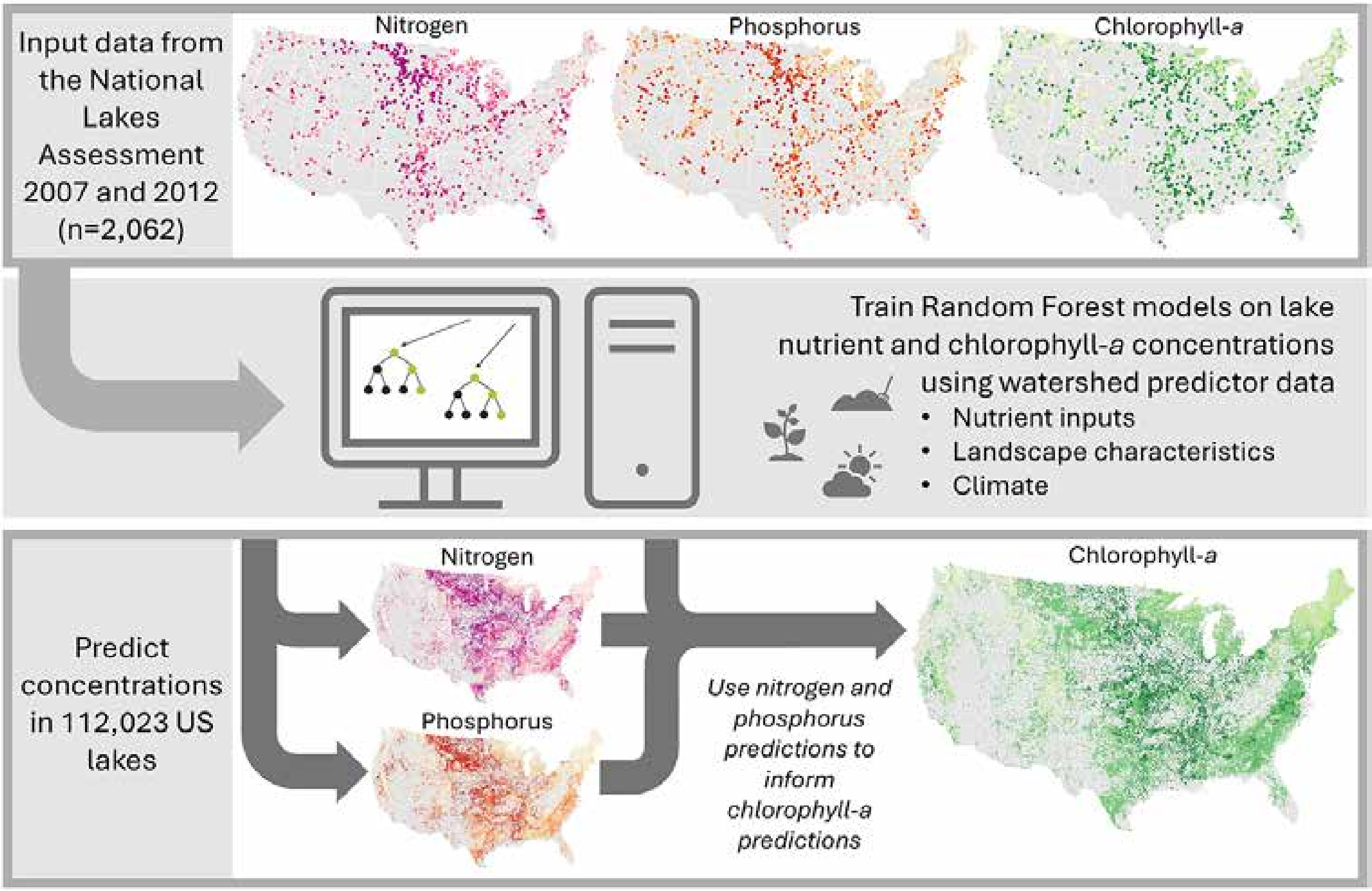

*Figure 1. Diagram showing the workflow that led to predictions of total nitrogen, total phosphorus, and chlorophyll-*a *concentrations for 112,023 US lakes.*

databases serve as one practical step forward for lake management and an example of what can be achieved when powerful open data and modern analytics come together.

References

Brehob, M.M., M.J. Pennino, A.M. Handler, J.F. Compton, S.S. Lee, andR.D. Sabo. 2024. Estimates of lake nitrogen, phosphorus, and chlorophyll-I concentrations to characterize harmful algal bloom risk across the United States. *Earth's Future*12(8): e2024EF004493.

Meredith Brehob is an aquatic ecologist with an ORISE fellowship at the U.S. EPA's Office of Research and Development. In this role, she works on applying national data to assess nutrient impacts on freshwater ecosystems and drinking water systems. Meredith has a Master of Science degree in Natural Resources and Environmental Science from the University of Nevada, Reno. To reach out, email Brehob.Meredith@epa.gov.

Robert Daniel Sabo is a scientist at the EPA's Office of Water. He began his career as an ORISE fellow with ORD in 2015 and joined the agency as a federal researcher in December 2019. Robert is the research lead for EPA's National Nutrient Inventory Research Portfolio, a multi-agency effort developing novel tools, models, and datasets designed to track the evolution of point and nonpoint sources of nutrient pollution to waterways across the U.S. and how changing environmental conditions impact waterbodies. Beyond research, Robert enjoys escaping to the mountain and Piedmont forests of Virginia for hiking and flyfishing with his wife and five daughters.

Michael Pennino is an ecologist with EPA's Office of Applied Sciences and Environmental Solutions in Washington, DC. He focuses his research on understanding spatial and temporal trends for indicators of environmental quality and human health, at regional and national scales. Michael uses using machine learning models and other data science techniques to help answer questions about our Nation's important water resources. Currently, Michael is leading projects that assess impacts of harmful algal blooms, urban best management practices, and other factors that influence contaminants in surface waters and drinking water. He holds a B.A. from Oberlin College (2005) and a Ph.D. from the University of Maryland, Baltimore County (2014).