UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
NATIONAL HEALTH AND ENVIRONMENTAL EFFECTS
RESEARCH LABORATORY
ATLANTIC ECOLOGY DIVISION
27 TARZWELL DRIVE • NARRAGANSETT, RI 02882

OFFICE OF
RESEARCH AND DEVELOPMENT

April 07, 2015

Pamela Silver, Ph.D.
Editor, *Freshwater Science*
School of Science
Penn State Erie
Erie, PA 16563

Dear Dr. Silver,

Please accept for your review the submission, "Modelling Lake Trophic State: A Random Forest Approach" to be published in *Freshwater Science*. In this article we detail our use of Random Forests to model trophic state in lakes across the United States. This research makes three important contributions. First, most prior studies of lake trophic state focus on a limited number of lakes that span local or regional extents. Our models are built for lakes that span the conterminous United States. Second, we compare models with and without *in situ* water quality data and find that models with only broadly available data still provide reliable predictions. This allows for predictions to be made in nearly every lakes in the U.S. Third, we utilize a data mining algorithm, Random Forests, to build our model. This approach provides accurate and unbiased predictions.

Our primary data sources are the US EPA's National Lakes Assessment from 2007, the National Land Cover Dataset from 2006, and predicted lake morphometry, which relies on the National Elevation Dataste and the National Hydrography Dataset Plus. These datasets are provided by the US Government, have all undergone strict and extensive quality assurance and are freely available. There are numerous publications based on these data, including some of our own prior work, most notably Hollister and Milstead (2010), Using GIS to Estimate Lake Volume from Limited Data, Lake and Reservoir Management 26(3), Hollister et al. (2011) Predicting Maximum Lake Depth from Surrounding Topography, PLoS ONE 6(9), and Milstead et al. (2013) Estimating Summer Nutrient Concentrations in Northeastern Lakes from SPARROW Load Predictions and Modeled Hydraulic Residence Time, PLoS ONE 8(11).

Lastly, we use open source tools, open data, and open access publishing whenever possible. Towards that end all of the data and source code for this manuscript are availble as an R package via GitHub (https://github.com/USEPA/LakeTrophicModelling). We also plan to take advantage of the Open Access publishing options provided by *Freshwater Science* and, if accepted, publish this manuscript as Gold Open Access. We are aware of both the additional cost for this and the page charges required by *Freshwater Science* and are able to cover these costs. We look forward to the review process and are hopeful that our manuscript can find a home in *Freshwater Science*.

Sincerely,

Jeffrey W. Hollister 401 782 9655 hollister.jeff@epa.gov

W. Bryan Milstead milstead.bryan@epa.gov

Betty J. Kreakie kreakie.betty@epa.gov