



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

September 03, 2015

Debra Peters
Editor, *Ecosphere*
USDA ARS
Las Cruces, NM 88003

Dear Dr. Peters,

Please accept for your review the submission, "Modelling Lake Trophic State: A Random Forest Approach" to be published in *Ecosphere*. In this article we detail our use of Random Forests to model chlorophyll *a* and 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. Lastly, we are proponents of practicing open science and as such, 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 available as an R package via GitHub (<https://github.com/USEPA/LakeTrophicModelling>) and a pre-print of this manuscript has been submitted to the PeerJ Preprint server (<https://peerj.com/preprints/1319/>). We understand that according to ESA policy this does not count as prior publication.

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