Deeper by the Dozen: Diving into a Database of 17,675 Depths for U.S. Lakes and Reservoirs

Katherine E. Webster, Ian M. McCullough, and Patricia A. Soranno

Abstract

Depth is a fundamental property of lakes, essential for understanding and quantifying a range of biological, chemical, and physical processes of individual lakes as well as the collective functioning of the millions of lakes across the globe. Despite this importance, lake depth is rarely available for large numbers of lakes across the broad scales of regions and continents. We describe a new open-access dataset, LAGOS-US DEPTH, which contains 17,675 maximum depths and 6,137 mean depths for lakes in the conterminous United States. These data represent one of the largest compilations of lake depths in the world and are connected to other data products through the LAGOS-US research platform. Here, we describe characteristics of lake maximum depth across the conterminous United States and identify gaps in data coverage that we encourage other researchers to fill by linking their own depth datasets for lakes within the LAGOS-US spatial footprint. Extending this open-access effort to include other regions and countries across the globe to build a more comprehensive and representative database of lake depth would better position scientists to quantify and articulate the critical roles that lakes play globally in biogeochemical cycling, maintaining biodiversity, and in maintaining the many ecosystem services that lakes provide.

Getting our feet wet

The depth of a lake is one of its most fundamental properties informing scientists on the biology, chemistry, and physical properties of lakes across the world, as well as anglers searching for the best place to fish. When scientists study individual lakes, lake depth is probably one of the first pieces of information acquired, and is relatively straightforward to measure. However, freshwater scientists are increasingly studying hundreds to millions of lakes across regions, continents, and the globe to examine such pressing issues as the estimated contribution of lakes to global carbon cycles or the projected trajectory of lake eutrophication and subsequent effects on the ecosystem services that lakes provide. Although lake depth is known for the world's largest (and often deepest) lakes, these lakes represent a tiny fraction of lakes globally. One might think that lake depth could be estimated from the depth of nearby lakes, from a lake's origin (e.g., glacial vs. volcanic), or from the nearby terrain of the land. However, none of these factors has been shown to accurately predict lake depth; in fact, models designed to predict lake depth from these types of factors have generated lake depth predictions with very large uncertainties that do little to improve our understanding of lake depth or to inform the use of lake depth to predict other important factors (Oliver et al. 2016; Stachelek et al. In press). Thus, we are left with having to measure lake depth directly rather than predicting it from easily acquired data sources. Perhaps ironically, models can more easily predict lake ecosystem properties such as water temperature and water clarity than depth itself, but depth is often a key predictor variable that limits the scope and accuracy of predictions (McCullough et al. 2012; Willard et al. 2021).

One reason for the lack of comprehensive lake depth data is that current and emerging technologies do not allow for the remote measurement of lake depth using satellite, aerial, or drone sensors; thus, lake depth can only be quantified by going out to the lake and taking measurements in a boat, which can be labor-intensive and requires direct access to the lake that is not always available. In addition, when lake depth measurements are taken, they are often done on an individual lake or on a regional basis by government agencies, researchers, private entities such as fishing groups, or individual lake property owners. Thus, the data that do exist are often stored in disparate places and not always easily accessible. Therefore, to increase the availability of lake depth data for scientists (and lake users) at broader spatial scales, datasets need to be compiled from a variety of different sources, an effort now more feasible thanks to web-based access to data and increasing open data practices by scientists and government agencies.

In this article, we describe our approach to compile lake depth data into a national-scale, publicly accessible database, LAGOS-US DEPTH, with maximum and mean depth data for, respectively, 17,675 and 6,137 lakes and reservoirs across the lower 48 U.S. states (Stachelek et al. 2021). We believe this is the largest compilation of maximum lake depth, particularly across this large of a geographic extent. We chose this footprint to match LAGOS-US LOCUS, another open-access dataset of the location, identifiers, and physical characteristics of 479,950 lakes greater than 1 ha in surface area and their watersheds in the conterminous United States that our research group has compiled (Cheruvelil et al. 2021; Smith et al. 2021). Our dataset provides maximum depth data on 3.7% of the lakes in this footprint, which is still a small fraction, but far better than any other available source of lake depths at this broad of a spatial extent. We use principles of open science to make these data available (Soranno et al. 2015), so any researcher can use our data to study this important lake property at an unprecedented spatial scale. We also hope this work will encourage scientists in other regions of the world to do the same, particularly for smaller lakes often overlooked in global datasets. Until new technologies emerge for the remote measurement of lake depth, these types of coordinated and open access compilation efforts will be the only way to improve continental and global estimates of important lake properties that rely on lake depth data, such as carbon and eutrophication, to name a few.

Uncovering hidden depths: the LAGOS-US DEPTH database

Our database contains depth data for an unprecedented number of lakes compiled manually from web data sources (e.g., state agency websites, fishing map databases). To focus our search, we first selected those lakes of the 479,950 LAGOS-US LOCUS lake population that had a known water quality sample taken at any point in the last 30 years. Then, for each lake in this subset, we searched manually for lake depth data. As a result, our sample selection was not random but is based largely on the population of lakes sampled for water quality. We describe more detailed methods in the LAGOS-US DEPTH User Guide that is available on EDI along with the database and associated metadata (Stachelek et al. 2021). For fuller use of this database, we recommend downloading and connecting to our research team's recent data product, LAGOS-US LOCUS (Cheruvelil et al. 2021; Smith et al. 2021), which enables the linking of these lake depth data to source datasets via the common unique lake identifier, lagoslakeid, includes Geographic Information System (GIS) layers with lake polygons, and provides data on geometry, location, and other lake properties.

The depth dataset includes data for both lakes and reservoirs representing a wide range of environmental conditions across the conterminous United States and across a broad surface area range. Here, we focus on maximum depth to describe some highlights of the patterns of lake depth across broad scales. In addition, we give an example of how connecting LAGOS-US DEPTH with other LAGOS-US data products (in this case a dataset on reservoir classification) provides a powerful source of data for addressing a wide variety of research questions. For example, these data would serve as a useful training dataset for future models predicting depth for the remaining 96.3% lakes in the conterminous United States that are not included in this data product.

Getting to the bottom of lake depth data availability

We were not surprised to find that the lakes that have been sampled for water quality and also have a lake depth measure are skewed towards larger lakes (Fig. 1), a bias previously documented by Stanley et al. (2019). The median surface area of lakes with depth values is ~ 50 ha, vs. the median area of all lakes in LAGOS-US LOCUS which is ~ 5 ha. Nevertheless, the lakes with depth values ranged in surface area from 1 to 338,950 ha, a span nearly extending to the full lake area range in LAGOS-US. Such a broad gradient will be valuable for future use of the dataset for building predictive models of lake depth across all size classes. Furthermore, above the size range of ~ 100 ha, we have a very good representation of lake depth values. Our dataset is actually quite complete for larger lakes and less so for moderately sized lakes; small lakes are very undersampled.

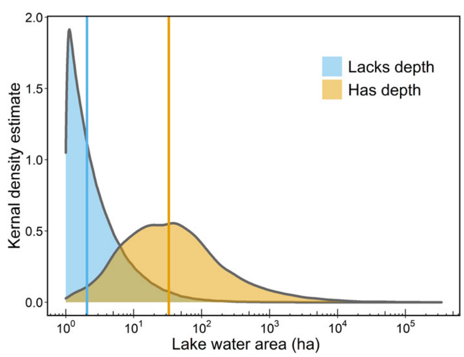

FIG. 1. The kernal density distribution of maximum depth values by lake surface area for lakes lacking maximum depth (blue color) and for lakes with maximum depth values (sand color). Vertical lines indicate median values for each group.

Regional differences in the number and percentage of lakes with lake depth data were related to the variation among U.S. states in both the absolute number of lakes and the availability of depth data (Fig. 2). For example, lake-dense Minnesota has the greatest number of lakes with maximum depth of any state, but Maine and New Hampshire have the greatest percent of their lake populations with lake maximum depth at 31%. All of these states have very accessible data for their lakes (e.g., http://www.dnr.state.mn.us/lakefind/ for Minnesota, and https://www.maine.gov/dep/

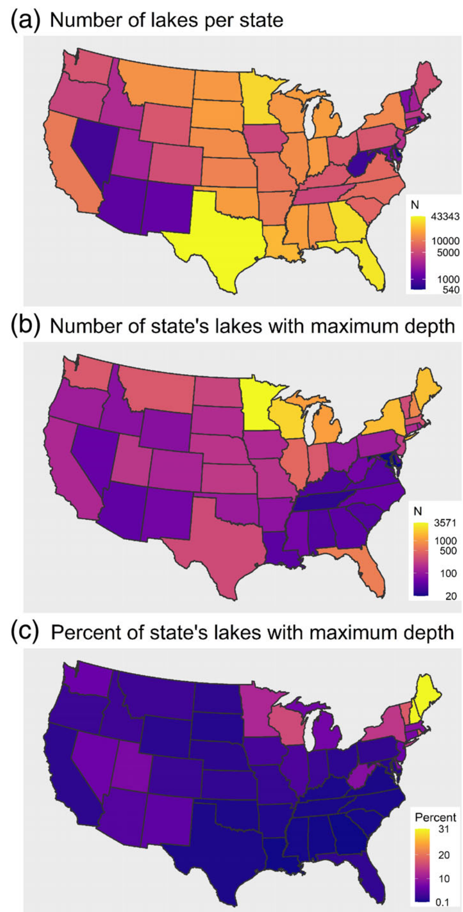

FIG. 2. Maps of the 48 U.S. states that depict (**a**) the number of LAGOS-US lakes in each state (plotted on log_{10} scale); (**b**) the number of lakes with maximum depth values in each state (plotted on log_{10} scale); and (**c**) the percent of LAGOS-US lakes in a state with maximum depth values in the dataset.

water/monitoring/lake/index.html for Maine), making data searching straightforward. Even with these excellent lake depth resources, we did not include all available lake depths from these sources if we did not also have a water quality value for the lake because the most labor-intensive part of our process was to connect the lake identifiers and location in our LAGOS-US LOCUS database, which is based on the National Hydrography Dataset (NHD), with the lakes and identifiers in these state databases, which we did one lake at a time. As more agencies and researchers develop crosswalk tables between such national-scale sources as the NHD and their state-based unique identifiers, this data integration will be quicker. For example, LAGOS-US LOCUS provides such a crosswalk table between the NHD unique identifiers and our unique identifier lagoslakeid. We encourage researchers interested in expanding on our database to fill in regional gaps in data by taking a similar approach. Such connections were beyond the scope of our study and are the reason why we provide all of our data and methods so that others can build on this first step toward compiling this important lake variable across broad spatial extents in the United States.

Dam it, what about reservoirs?

LAGOS-US DEPTH contains more maximum depth data for natural lakes (57%) than reservoirs (43%), which we discovered after linking our data with LAGOS-US-RESERVOIRS, a classification dataset of U.S. lakes ≥ 4 ha that differentiates natural lakes and reservoirs (Polus et al. 2021). This pattern is consistent across waterbody size classes (Fig. 3a), even though reservoirs represent ∼ 44% of waterbodies ≥4 ha across the conterminous United States as defined by Polus et al. (2021). Interestingly, the distribution of maximum depths does not differ dramatically between natural lakes and reservoirs and median values for both are ∼ 7 m (Fig. 3b). Although, as described above, we have more complete depth data for larger waterbodies in general, we have depth data for the majority of natural lakes ≥ 100 ha but only for the majority of reservoirs ≥250 ha. Notably, reservoirs are considered more sensitive to depth fluctuations due to water withdrawals and/or climatic variability; if so, these single-point estimates of depth are likely more variable for reservoirs than for natural lakes. Ecological and hydrological differences between natural lakes and reservoirs notwithstanding, researchers should recognize that although LAGOS-US DEPTH generally contains more depth data for natural lakes, reservoirs are still well represented.

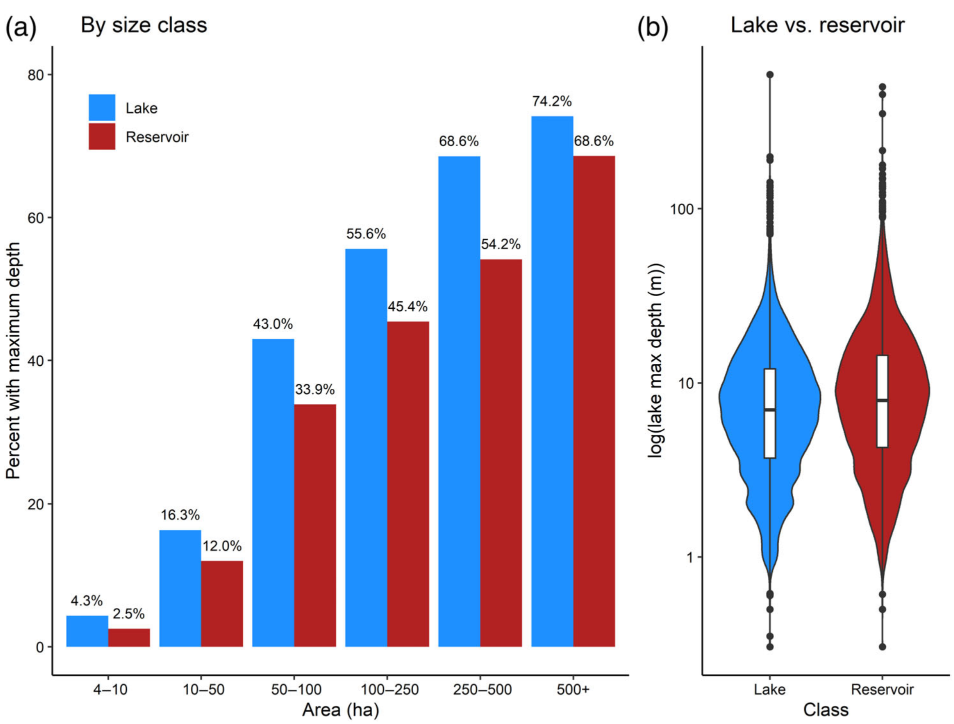

FIG. 3. Representation of maximum depth data across natural lakes and reservoirs by lake area size class (**a**) and distributions of maximum depth of lakes classified as (natural) lakes and reservoirs on a $\log_{10}$ scale (**b**). Percentages indicate percent of waterbodies in each size class and classification type with maximum depth data in our dataset. Lakes < 4 ha could not be reliably classified as natural lakes or reservoirs and are thus not shown.

Deep pockets: enriching our understanding of depth spatial patterns

Across our dataset, maximum depth ranges from <1 to 589 m, with a median of 7 m. The deepest lakes in the conterminous United States from our database are mostly in the western United States, particularly in California, Oregon, Washington, and the Rocky Mountain states (Fig. 4). Based on averaged depths, California, Oregon, and Wyoming are ranked 1–3 in terms of statewide mean maximum lake depth with Colorado (5), Washington (6), and Idaho (10) also ranked in the top 10. Interestingly, the other 4 states in the top 10 are in the southeastern United States: South Carolina (4), Tennessee (7), Alabama (8), and North Carolina (9). Notably, however, there are pockets of especially deep lakes in central New York, the Southern Appalachians, Minnesota, and Wisconsin. Most of the Great Plains region and southeastern United States are dominated by shallower lakes (< 10 m), whereas formerly glaciated New England and the Upper Midwest have many more moderately deep lakes in the 15–50 m range. These data demonstrate regional concentrations of very deep lakes, as well as pockets of individually deep lakes in regions dominated by shallower lakes.

However, this basic information requires some deeper context (this is a depth article, after all). Most of the deepest lakes in the conterminous United States are classified as reservoirs. There are 46 lakes ≥ 100 m deep, 31 and 15 of which are reservoirs and natural lakes, respectively. Furthermore, a subset of these very deep "natural" lakes originated as mining pits and therefore are not true natural lakes (see Polus et al. 2021 for definitions). Of the 10 deepest U.S. lakes, 7 are reservoirs and just 3 are natural lakes, although the deepest lake overall is Crater

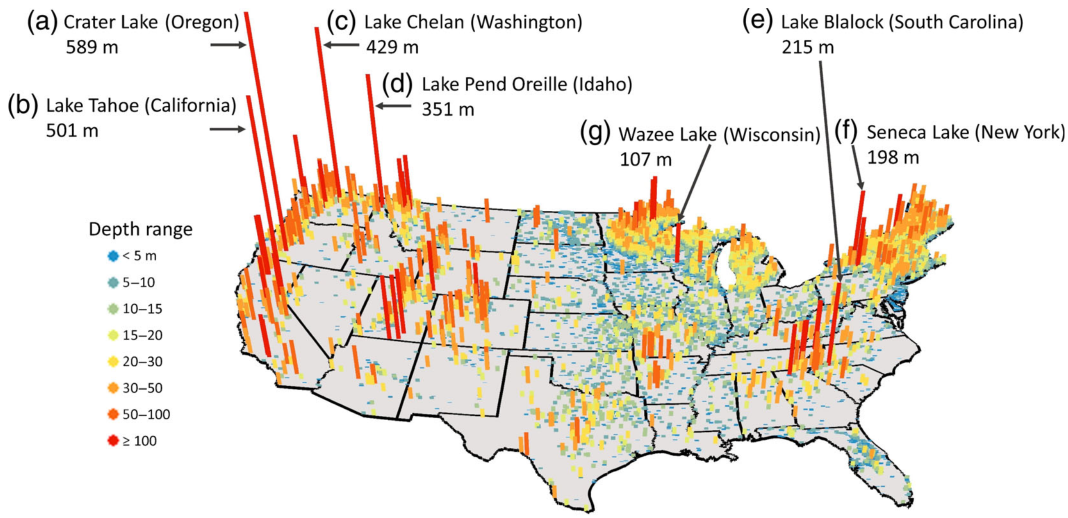

FIG. 4. Map depicting maximum depths for all lakes in the dataset in bins of lake depth ranges with blues and greens being the shallowest depths and reds and oranges being the deepest. Labeled lakes include the six deepest lakes overall (**a**–**f**) and an unusually deep Midwestern lake formed from an abandoned mine (**g**).

Lake, Oregon (Fig. 4a), a natural lake of volcanic origin. Interestingly, only 2 of the top 10 deepest lakes are in the Eastern United States; one is a reservoir (Lake Blalock, South Carolina; Fig. 4e) and one is a natural lake (Seneca Lake, New York; Fig. 4f). In the Eastern United States, the deepest lakes are natural lakes in central New York and reservoirs in North Carolina, South Carolina, and Georgia. In the Midwestern United States, many of the deepest lakes are artificial lakes that originated as mining pits, for example, Wazee Lake in Wisconsin (Fig. 4g). Overall, the deepest lakes in the conterminous United States are mostly found in the Western United States (Fig. 4a–d) where there are fewer lakes and reservoirs.

As a final note of caution to users, we know there are many lakes with depth values that are not in our dataset and likely some lakes that are well-known at the state-level or beyond. Therefore, this dataset cannot be used as the definitive list of "deepest lakes in the United States". Rather, we present it as a snapshot of lake depth across the conterminous United States that can demonstrate broad regional patterns, but not show the extreme values (i.e., deepest lakes in a state) with great accuracy nor represent fluctuations in lake depth over time.

Next steps to move from surficial to profundal understanding

Our dataset of 17,675 lake maximum depths and 6,137 mean depths is, to our knowledge, the largest compilation of lake depths that also represents lakes that span broad ecological and geophysical gradients. Despite the relatively small proportion that this number represents compared to the full lake population, we found interesting spatial patterns that were difficult to detect prior to compiling the data. In fact, our descriptions of this dataset have just scratched the surface of potential knowledge that can be gained from examining lake depth at this broad spatial scale. Given the well-recognized importance of lake depth for understanding lakes and for quantifying their significance at regional, national, or global scales, we were somewhat surprised by the relative scarcity of easily accessible and well-documented lake depth data in the United States. With LAGOS-US DEPTH and the documentation that we provide, we encourage other researchers to fill in the gaps noted in our paper (e.g., small lakes and lakes without water quality data). We have designed all of the LAGOS-US data products to be extensible—in other words, designed to allow future additions by other researchers by sharing detailed documentation and description of our methods, GIS layers, unique identifiers, and a crosswalk table of common lake identifiers for the United States (Soranno et al. 2015; Cheruvelil et al. 2021; Smith et al. 2021). Distributing the compilation of such data across research groups would greatly facilitate continental and global-scaled studies, ensure continued open access, and recognize the contributions and efforts of the researchers compiling the data. We particularly encourage compilations for regions and countries outside of the footprint of this database. We know that national efforts do exist, such as Sweden's extensive dataset for lakes that includes lake depth measurements. We hope that our dataset and approach inspire other researchers to add to this effort by building regional or national lake depth modules that are themselves deposited in open-access data repositories, that recognize the efforts of local authors who have conducted the research to compile the data, and that can be linked to the LAGOS-US data products to create a global dataset of lake depth that represents a full range of lakes and reservoirs. We welcome any researcher to contact us for guidance on creating such data modules, to share the relatively straightforward steps needed to create open-access datasets (Soranno 2019), and to show the best ways to ensure compatibility with LAGOS-US as a starting point for open-access global datasets that are designed from the start to be added to by the research community.

Acknowledgments

We thank Jem Stachelek who led the development of this dataset and the individuals who helped to compile lake depth data and/or draft the User Guide that describes this dataset (Lauren K. Rodriguez, Jessica Díaz Vázquez, Arika Hawkins, Ellie Phillips, Allie Shoffner, Katelyn B.S. King, Jake Namovich, Lindsie A. Egedy, Maggie Haite, and Patrick

J. Hanly). Thanks to Maggie Haite for conducting quality control checks on some of the data and to Kendra S. Cheruvelil who leads the development of LAGOS-US data products. Laura Falkenberg and Chris Filstrup provided helpful review comments. This study was supported by U.S. National Science Foundation Macrosystems Biology Program grants, DEB-1638679; DEB-1638550, DEB-1638539, DEB-1638554. PAS was also supported by the USDA National Institute of Food and Agriculture, Hatch project 1013544.

Author contributions

All authors contributed equally to this article in addition to the following specific contributions. KEW made Figs. 1, 2, 4; IMM made Fig. 3. KEW integrated the datasets for analysis, IMM led the effort to QAQC the data in preparation for this article, and PAS had the initial idea for this article and co-leads the development of this and other LAGOS-US data products.

References

Cheruvelil, K. S., P. A. Soranno, I. M. McCullough, K. E. Webster, L. K. Rodriguez, and N. J. Smith. 2021. LAGOS-US LOCUS v1.0: Data module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. Limnol. Oceanogr. Lett. **6**: 270–292. doi:10.1002/lol2.10203.

McCullough, I. M., C. S. Loftin, and S. A. Sader. 2012. Combining lake and watershed characteristics with Landsat TM data for remote estimation of regional lake clarity. Remote Sens. Environ. **123**: 109–115.

Oliver, S. K., P. A. Soranno, E. Fergus, T. Wagner, L. A. Winslow, C. E. Scott, K. E. Webster, J. A. Downing, and E. H. Stanley. 2016. Prediction of lake depth across a 17-state region in the United States. Inland Waters **6**: 314–324. doi:10.5268/IW-6.3.957.

Polus, S.M., and others. 2021. LAGOS-US RESERVOIR: Data module classifying conterminous U.S. lakes 4 hectares and larger as natural lakes or reservoirs. Environmental Data Initiative ver 1. Environmental Data Initiative. 10.6073/pasta/c850e645d79bb239e1dfeadd0af6b631.

Smith, N.J., K.E. Webster, L.K. Rodriguez, K.S. Cheruvelil, and P.A. Soranno. 2021. LAGOS-US LOCUS v1.0: Data module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. ver 1. Environmental Data Initiative. 10.6073/pasta/e5c2fb8d77467d3f03de4667ac2173ca.

Soranno, P. A. 2019. Six simple steps to share your data when publishing research articles. Limnol. Oceanogr. Bull. **28**: 41–44. doi:10.1002/lob.10303.

Soranno, P. A., and others. 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: Fostering open science and data reuse. GigaScience **4**: 28. doi:10.1186/s13742-015-0067-4.

Stachelek, J., and others. 2021. LAGOS-US DEPTH v1.0: Data module of observed maximum and mean lake depths for a subset of lakes in the conterminous U.S. ver 1. Environmental Data Initiative. 10.6073/pasta/64ddc4d04661d9aef4bd702dc5d8984f.

Stachelek, J., P. J. Hanly, and P. A. Soranno. In press. Imperfect slope measurements drive overestimation in geometric cone model of lake and reservoir depth. Inland Waters. doi:10.1080/20442041.2021.2006553.

Stanley, E. H., S. M. Collins, N. R. Lottig, S. K. Oliver, K. E. Webster, K. S. Cheruvelil, and P. A. Soranno. 2019. Biases in lake water quality sampling and implications for macroscale research. Limnol. Oceanogr. **64**: 1572–1585. doi:10.1002/lno.11136.

Willard, J. D., J. S. Read, A. P. Appling, S. K. Oliver, X. Jia, and V. Kumar. 2021. Predicting water temperature dynamics of unmonitored lakes with meta-transfer learning. Water Resour. Res. **57**: e2021WR029579.

Katherine E. Webster, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA; katherine.e.webster@gmail.com

Ian M. McCullough, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

Patricia A. Soranno, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA