Hydrobiologia (2022) 849:3663–3677
https://doi.org/10.1007/s10750-022-04811-1

Taking a macroscale perspective to improve understanding of shallow lake total phosphorus and chlorophyll *a*

Kendra Spence Cheruvelil · Katherine E. Webster · Katelyn B. S. King · Autumn C. Poisson · Tyler Wagner

Received: 31 May 2021 / Revised: 20 January 2022 / Accepted: 29 January 2022 / Published online: 25 February 2022

Abstract We conducted a macroscale study of 2210 shallow lakes (mean depth ≤ 3 m or a maximum depth ≤ 5 m) in the Upper Midwestern and Northeastern USA. We asked the following: What are the patterns and drivers of shallow lake total phosphorus (TP), chlorophyll *a* (CHLa), and TP–CHLa relationships at the macroscale, how do these differ from those for 4360 non-shallow lakes, and do results differ by hydrologic connectivity class? Spatial patterns and Bayesian hierarchical models indicated that shallow lakes had higher TP and CHLa than non-shallow lakes, connected shallow lakes were more productive than unconnected shallow lakes, and there was regional variation in these patterns. Important predictors of TP and CHLa included lake-specific watershed:lake area ratio, forested land use/cover, and baseflow; unconnected lakes were more difficult to predict than connected lakes; and region-specific predictors were mostly unimportant. Shallow lake TP–CHLa relationships were less steep than for non-shallow lakes and these relationships varied regionally. Our results, combined with the facts that only 23% of lakes in the study extent have depth data and that shallow and unconnected lakes are undersampled, have important implications for estimates of lake contributions to global cycles that are based mainly on large (and deeper) lakes.

Kendra Spence Cheruvelil and Katherine Webster are co-first/lead authors

Guest editors: José L. Attayde, Renata F. Panosso, Vanessa Becker, Juliana D. Dias & Erik Jeppesen / Advances in the Ecology of Shallow Lakes

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10750-022-04811-1.

K. S. Cheruvelil (✉) · K. E. Webster · K. B. S. King · A. C. Poisson
Department of Fisheries and Wildlife, Michigan State University, 480 Wilson Rd, East Lansing, MI 48824, USA
e-mail: ksc@msu.edu

K. S. Cheruvelil
Lyman Briggs College, Michigan State University, 919. E Shaw Lane, East Lansing, MI 48825, USA

Present Address:
K. B. S. King
School for Environment and Sustainability, University of Michigan, 440 Church St, Ann Arbor, MI 48109, USA

Present Address:
A. C. Poisson
U.S. Geological Survey St. Petersburg Coastal and Marine Science Center, 600 4th Street South, Saint Petersburg, FL 33701, USA

T. Wagner
U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Unit, The Pennsylvania State University, 402 Forest Resources Building, University Park, PA 16802, USA

Keywords Phosphorus · Chlorophyll · TP:CHL · Shallow lakes · Macroscale · LAGOS

Introduction

Lake productivity, generally assessed as the concentration of pelagic chlorophyll *a* (CHLa) and modeled as a function of nutrients such as total phosphorus (TP), is fundamental to our understanding of lake ecosystems and their management. Numerous studies have examined controls on CHLa and TP concentrations, in particular regarding relationships with lake depth and watershed features such as land use and climate (e.g., Taranu & Gregory-Eaves, 2008; Read et al., 2015; Stachelek et al., 2020; Shuvo et al., 2021). Statistical modeling of the TP–CHLa relationship, initiated by Vollenweider's classic study (1975), has continued to the present and greatly contributed to the management of lakes. Recent efforts, such as Quinlan et al. (2021), have extended studies of individual lakes through time and comparative studies of lakes across regions to examine the role of hierarchical controls on lake states and relationships at the macroscale (i.e., regions to continents; Heffernan et al., 2014).

Such macroscale studies increase the understanding of populations of lakes and improve the ability to extrapolate results to unsampled lakes, scale-up local results to the globe, and forecast lake responses to change through time (Cheruvelil & Soranno, 2018). In fact, macroscale applications of hierarchical models have demonstrated that average TP and CHLa differ regionally, and that the drivers of regional TP–CHLa relationships can be non-linear, multi-scaled, and include cross-scale interactions (Jackson et al., 2007; Phillips et al., 2008; Wagner et al., 2011; Filstrup et al., 2014; Quinlan et al., 2021). Macroscale research has also demonstrated that lake TP and CHLa are related to lake hydrologic location, climate, and the level of connectivity to other waterbodies. For example, highly connected lakes in North America have higher average concentrations of TP and CHLa than lakes that are isolated from or form the headwaters of lake chains (Soranno et al., 1999; Martin & Soranno, 2006; Zhang et al., 2012).

To our knowledge, there has been less research about the macroscale patterns of shallow lake TP and CHLa concentrations and of TP–CHLa relationships. Past local to regional-scale shallow lake research highlights some key differences between shallow lakes and lakes more generally. For example, shallow lake productivity is strongly affected by littoral and benthic processes and interactions (Vander Zanden & Vadeboncoeur, 2020), with the relative importance of these processes related to lake depth. Shallow lakes often do not thermally stratify, have high potential for nutrient resuspension from sediments, and have a large proportion of lake volume receiving light penetration sufficient for primary production (e.g., Scheffer, 1998; Kalff, 2001), which are factors that influence TP and CHLa concentrations, as well as the TP–CHLa relationship (Brett & Benjamin, 2008). Further, primary production is dominated by macrophytes and benthic algae when shallow lakes are clear, thus affecting the relationship between TP and CHLa (Scheffer, 1998; Scheffer et al., 1993). The effects of surface water connectivity on shallow lake TP and CHLa at the macroscale are not well understood. As a result, there exists a knowledge gap of how shallow lake TP, CHLa, and TP–CHLa relationships vary across regions, what drives these states and relationships, and whether they differ by surface water connectivity.

Newly compiled and harmonized macroscale datasets about lakes and their ecological settings provide scientists with an opportunity to address these knowledge gaps. We used LAGOS-NE (Soranno et al., 2015, 2017), which includes thousands of lakes in the Northeastern and Upper Midwestern USA where there are wide spatial gradients in many ecological setting variables (Collins et al., 2017; Lapierre et al., 2018; Stachelek et al., 2020). We used these data to ask the following: What are the patterns and drivers of shallow lake TP, CHLa, and TP–CHLa relationships at the macroscale? Do the patterns and drivers of shallow lakes differ from those for non-shallow lakes and do they differ by lake connectivity class? Based on previous research demonstrating the relative importance of lake depth and surface water connectivity for driving lake productivity, we expected lake TP and CHLa to be highest and TP–CHLa relationships to be steepest (i.e., larger slope parameter estimates) for connected and shallow lakes. Prior macroscale lake research, irrespective of depth, also led us to expect shallow lake TP, CHLa, and TP–CHLa relationships to vary regionally and both fine- and broad-scale landscape features (and interactions between them) to drive those patterns. Our research can help limnologists understand the patterns and drivers, particularly connectivity, of shallow lakes at the macroscale, which is important during this time of rapid global changes.

Methods

Study lakes and ecological setting

Lake TP, CHLa, and geospatial landscape data came from the LAGOS-NE database. LAGOS-NE is a database that extends approximately 1,800,000 km^2 in a 17-state region of the Upper Midwestern and Northeastern United States (Soranno et al., 2015, 2017). This database comprehensively describes the landscape, climatic, and surface water context for 51,101 lakes ≥ 4 ha in surface area. Approximately 25% of the LAGOS-NE lakes have at least one observation of water quality data compiled from 87 disparate datasets and harmonized and documented for regional-scale limnological study (Soranno et al., 2015). We supplemented lake depth data in LAGOS-NE with a new dataset of compiled lake depth data for 17,675 lakes in conterminous USA (LAGOS-US: DEPTH; Stachelek et al., 2021). LAGOS-NE data used for this shallow lake research are lake water chemistry from $\text{LAGOS}_{\text{LIMNO}}$ version 1.087.3 and geospatial data from $\text{LAGOS}_{\text{GEO}}$ version 1.03 (Soranno et al., 2017) and LAGOSclimate (Collins et al., 2019).

To select our study lakes from LAGOS-NE, we included lakes with either mean or maximum depth data ($n = 11{,}801$). The study lakes were further restricted to include those with TP or CHLa data from either surface or epilimnion water samples collected during the summer stratified season (June 15th–September 15th) of 2002–2011, and to lakes within the study extent defined by hydrologic subregion (4-digit hydrologic unit, HU4; Seaber et al., 1987) polygons defining the study area, resulting in 6570 study lakes (Fig. 1). Because we were interested in spatial variability in TP and CHLa, we used the median concentrations for each lake over the two-decade time period, resulting in a single value per lake. We defined 'Shallow' lakes as the subset of study lakes with a mean depth ≤ 3 m ($n = 903$ lakes) or, where mean depth data were lacking, a maximum depth ≤ 5 m ($n = 1307$ lakes) (sensu Scheffer, 1998; Jeppesen et al., 1997; Padisák & Reynolds, 2003). This process resulted in 2210 Shallow lakes with either TP or CHLa values and wide ranges of characteristics (Table 1). The remaining 4360 'non-Shallow' lakes were analyzed separately for comparative purposes. This subset included lakes with mean depths just above the shallow lake cutoff to very deep (mean depth range = 3.0 to 88.6 m; Table 1).

We quantified each lake's landscape setting using geospatial variables (Supplement 1) that characterize sources, transport, and internal processing of nutrients (Collins et al., 2017). The geospatial predictor variables were derived at two spatial scales that characterize local and regional drivers of lake TP and CHLa. The 24 local-scale variables were derived for the individual lake, the lake's watershed, or, in the case of climate and hydrology variables, for each lake's hydrologic watershed (12-digit hydrologic unit, HU12; Seaber et al., 1987). Because the median and 75th quantile of the number of lakes contained within a HU12 was 1 (maximum of 29), most lakes had a unique value for all HU12 predictors. Ten regional-scale predictor variables were calculated at the HU4 scale (number of lakes per HU4 ranged from 1 to 1133, median = 61; total number of HU4 was 63; Supplement 1). Lakes were also classified into two categories of surface water connectivity. First, 'unconnected' lakes include isolated and headwater lakes. Isolated lakes have no inlet or outlet stream connections and headwater lakes have no inlet but have an outflowing stream connection ($n = 1967$). Second, 'connected' lakes include lakes with any inflowing streams, regardless of number or whether there were also upstream lakes ($n = 4603$).

Statistical modeling

Identifying drivers of lake TP and CHL

We used hierarchical Bayesian linear models to simultaneously identify important local- and regional-scale geospatial drivers of TP and CHLa. The hierarchical model was as follows:

$$y_i \sim N\left(\alpha_{j(i)} + \beta_1 X_{1i} + \cdots + \beta_p X_{pi}, \sigma^2\right) \text{ for } i = 1, \ldots n$$

$$\alpha_j \sim N\left(\delta_0 + \gamma_1 Z_{1j} + \cdots + \gamma_m Z_{mj}, \sigma_\alpha^2\right) \text{ for } j = 1, \ldots J$$

where y_i is natural log-transformed TP or CHLa for lake i, α_j is the region-specific intercept for HU4 j,

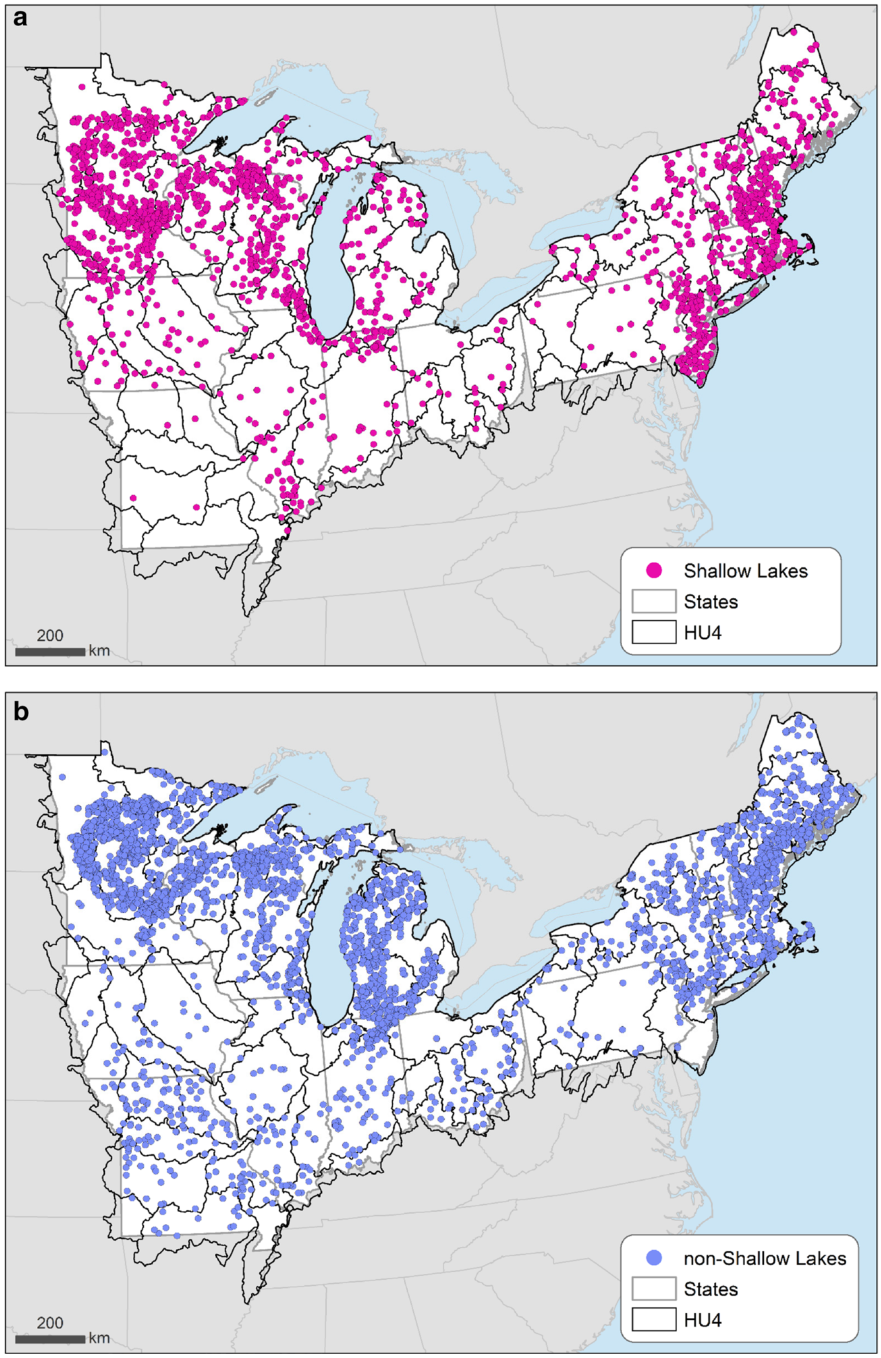


Fig. 1 Location of the 6570 study lakes subset by depth. Study lakes are those LAGOS-NE lakes having data on mean or maximum depth and CHLa and/or TP concentrations. **A** Shallow lakes are those with a mean depth ≤ 3 m (n = 903 lakes) or, where mean depth data were lacking, a maximum depth ≤ 5 m (n = 1307 lakes) and **B** non-Shallow lakes are all other study lakes. HU4s are hydrologic sub-regions (4-digit hydrologic unit, HU4; Seaber et al., 1987)

$\beta_1 \ldots \beta_p$ are the estimated effects of local-scale predictors $X_1 \ldots X_p$, σ^2 is the residual variance, δ_0 is the fixed intercept, $\gamma_1 \ldots \gamma_m$ are the estimated effects of region-scale predictors $Z_1 \ldots Z_m$, and σ_α^2 is the conditional among region variance. To accommodate the fact that there were many local and regional-scale predictors, we fitted parameters using a Bayesian LASSO approach (least absolute shrinkage and selection operator; Tibshirani, 1996; Park & Casella, 2008), which is not particularly sensitive to collinearity (Dormann et al., 2013). Because we divided the predictor variables into two groups (local- and

Table 1 Descriptive statistics of the study lakes, subset by depth (i.e., Shallow and non-Shallow)

	n	Median	Mean	Min	Max	25th	75th
Shallow Lakes ($n=2210$)							
Mean depth (m)	903	2.1	2.0	0.3	3.0	1.5	2.5
Max depth (m)	2190	3.4	3.9	0.3	32.3*	2.2	4.9
Lake area (ha)	2210	43	126	4	6905	15	100
TP (µg/l)	1394	24.0	58.0	3.0	1122.5	14.0	59.0
CHLa (µg/l)	1955	10.5	28.5	0.3	403.0	4.4	37.0
Non-Shallow Lakes ($n=4360$)							
Mean depth (m)	1513	5.2	6.2	3.0	88.6	4.0	7.1
Max depth (m)	4354	12.2	14.4	3.7	198.4	8.5	17.4
Lake area (ha)	4359	74	429	4	123,780	33	190.5
TP (µg/l)	2786	13.5	24.1	2.0	600.0	8.5	24.5
CHLa (µg/l)	3488	5.1	12.3	0.2	198.2	2.9	12.4

Lakes are ≥ 4 ha in surface area, and have mean or maximum depth information as well as summer TP and/or CHLa data collected between 2002 and 2011. Shown are the total number of lakes (n), median, mean, minimum (Min), maximum (Max), and 25th and 75th percentiles

*For the 903 lakes classified as shallow using mean depth ≤ 3 m, there were 391 lakes with a maximum depth > 5 m

regional-scale), each group received a separate double exponential (DE) LASSO prior such that $\beta_p \sim DE(0, \lambda_1)$, and $\gamma_m \sim DE(0, \lambda_2)$. A diffuse normal prior was used for δ_0, diffuse uniform priors were used for σ and σ_α, and λ_1 and λ_2 were given diffuse gamma priors. Models were fitted for each response variable (TP, CHLa), lake connectivity class (connected, unconnected, both classes combined), and depth subset (Shallow, non-Shallow) combination separately. All predictor variables were transformed (natural log or logit-transformed in the case of proportions) and standardized (mean = 0, SD 1) prior to analysis.

Identifying cross-scale interactions affecting the TP–CHLa relationship

A hierarchical Bayesian model was used to quantify potential cross-scale interactions (CSIs; processes operating at one spatial or temporal scale interact with processes operating at another scale; Heffernan et al., 2014; Soranno et al., 2014) in the TP–CHLa relationships. The model was a varying intercept, varying slope model with a single lake-scale predictor (i.e., $\log_e$[TP]) and regional predictors used to model the variability in the slopes describing the TP–CHLa relationships across regions (Wagner et al., 2016). The hierarchical model was as follows:

$$y_i \sim N\left(\alpha_{j(i)} + \beta_{j(i)} X_i, \sigma^2\right) \text{ for } i = 1, \ldots n$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim MVN\left(\begin{pmatrix} \gamma_0^\alpha \\ \gamma_0^\beta + \gamma_1^\beta Z_{1j} + \gamma_m^\beta Z_m \end{pmatrix}, \Sigma\right) \text{ for } j = 1 \ldots J$$

where y_i is natural log-transformed CHLa for lake i, X_i is natural log-transformed TP for lake i, α_j and β_j are intercepts and slopes, respectively, that describe the TP–CHLa relationship for HU4 j, γ_0^α is the grand mean intercept (across all lakes), γ_0^β and γ_m^β are the intercepts and slopes describing the relationships between regional-scale predictors and the slopes in the relationship between TP–CHLa (γ_m^β represent the estimated CSIs). Diffuse normal priors were used for γ_0^α and γ_0^β, a LASSO prior was used for γ_m^β, and we modeled the variance–covariance matrix (Σ) using the scaled inverse-Wishart distribution (Gelman & Hill, 2007). Because lake connectivity is a categorical variable and we had a priori reasons to believe that these classes affect lake states and relationships, models were fitted for each lake connectivity class separately. All continuous predictor variables were transformed (natural log or logit-transformed) and standardized (mean = 0, SD 1) prior to analysis.

Identifying drivers and CSIs

For both sets of models described above, we ran three parallel Markov chains beginning each chain with random starting values. Each chain was run for 60,000 iterations with the first 50,000 iterations discarded as burn-in. This resulted in a total of 30,000 samples used to summarize the posterior distributions of the model parameters. Convergence was assessed visually through inspection of trace plots and quantitatively using the Brooks-Gelman-Rubin statistic (Brooks & Gelman, 1998). To assess whether intercepts of the TP or CHLa predictive models or intercepts or slopes of the TP—CHLa models differed between Shallow and non-Shallow study lakes within each connectivity class (connected, unconnected, and both), we calculated the posterior difference for each paired comparison and then determined whether or not the 95% credible interval of the difference overlapped with zero. Model intercepts or slopes were considered different when credible intervals did not overlap zero. Marginal (fixed-effects only; R^2_{marg}) and conditional R^2 values (fixed + random effects; R^2_{cond}) were calculated for predictive models for TP and CHLa (Nakagawa & Schielzeth, 2013). Models were fitted by calling the program JAGS (Plummer, 2003) using the jagsUI package (Kellner, 2019) in the program R (R Core Team, 2021).

Results

Compared to the LAGOS-NE census lake population, the study lakes, Shallow lakes, and non-Shallow lakes were overly represented by larger lakes (Fig. 2). However, shallow lakes tended to be smaller in area and thus slightly more representative of the area distribution of the entire census population compared to both the study lakes (Table 1) and non-Shallow lakes (Fig. 2).

The study lakes were densest across the northern portion of the study extent, with both Shallow and non-Shallow lakes most common along the eastern USA border and the northwestern parts of the study extent (Fig. 3). Notably, there were fewer Shallow compared to non-Shallow lakes with TP or CHLa data along the southern edge of the extent, especially in the southwestern portion of our study extent (Fig. 1). Both connected and unconnected

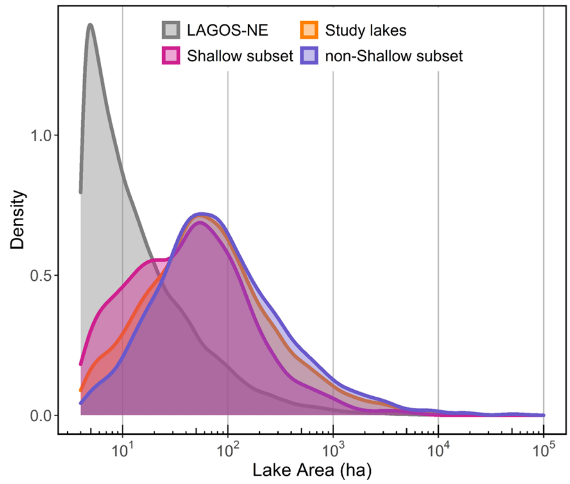


Fig. 2 The density distribution of lake depth by lake surface area for the census population of LAGOS-NE lakes; all study lakes (those with depth and either TP or CHLa); and all study lakes subset by depth into Shallow and non-Shallow lakes

lakes showed increasing density gradients from south to north. Connected lakes (70% of the study lakes) occurred at higher densities than unconnected lakes, and unconnected shallow lakes were particularly rare in southern parts of the study extent (Fig. 3).

Lake TP and CHLa ranged several orders of magnitude (Fig. 4A, C) with TP and CHLa both higher in Shallow lakes compared to non-Shallow lakes, and in connected lakes as compared to unconnected lakes (Fig. 4A, C). Based on criteria in Dodds et al. (2006), the interquartile range of trophic status based on TP was mainly mesotrophic. However, CHLa-based trophic status extended from the mesotrophic to eutrophic range. Shallow lakes tended to have higher distributions of both TP and CHLa than non-Shallow lakes, while connected lakes tended to have higher concentrations than did unconnected lakes. These patterns were stronger for TP than for CHLa (Fig. 4).

TP and CHLa exhibited a spatial pattern of increasing concentration from north to south, with HU4 medians of trophic state indicating more oligotrophic conditions in the north central and northeastern regions, and hyper-eutrophic regions only in the southwest corner of the study extent (Fig. 4B, D). With both connectivity classes combined, regional (HU4) median trophic status tipped into hyper-eutrophic status more often for Shallow lakes (e.g., 34% and 24% of HU4s, respectively, for CHLa

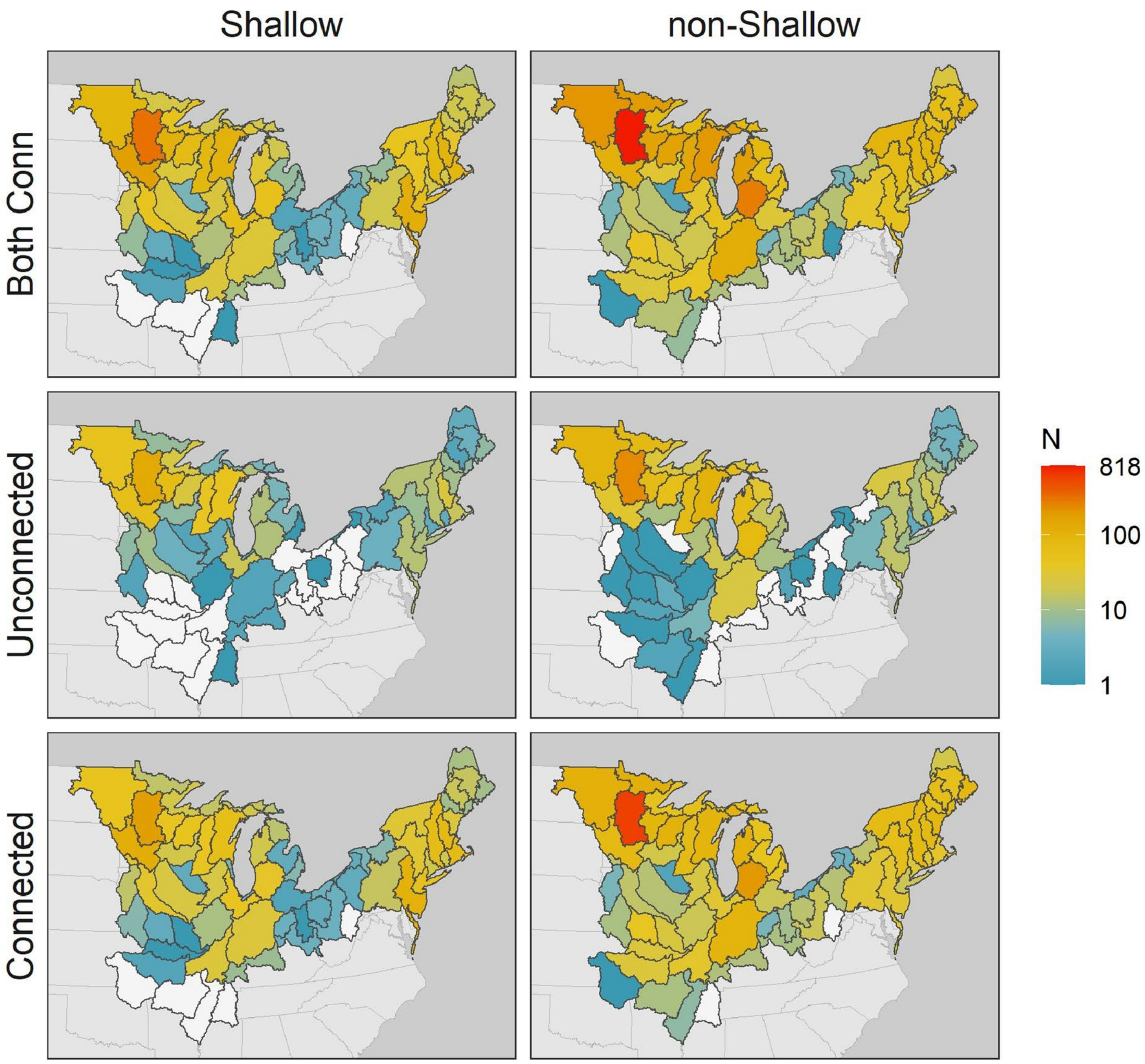


Fig. 3 Geographic distribution of study lakes (number per HU4) mapped by connectivity class (rows showing both connectivity types, unconnected lakes and connected lakes) for Shallow lakes and non-Shallow lakes (left and right columns). Unconnected lakes (middle row) lack stream inlets, while connected lakes (bottom row) have inlet streams that may connect them to upstream lakes; the top row shows all connectivity classes combined. Darker gray lines delineate HU4 regional boundaries; white shaded HU4s have no lakes representing the specific connectivity class by depth subset combination. Light gray lines delineate US states

and TP) than for non-Shallow lakes (14% and 0%; Fig. 4B, D, top row of panels).

The predictor variables we modeled exhibited broad ranges across the study extent, including both the lake-specific variables such as lake area and water residence time (as estimated by the ratio between watershed and lake area) and the ecological setting variables reflecting land use/cover, hydrology, and climate (Supplement 1). For example, lake area ranged over four orders of magnitude, while the percentage of urban land use and wetland cover in the study lake watersheds ranged from 0 to 100 and 0 to 93, respectively.

For both Shallow and non-Shallow lake subsets, hierarchical modeling results (Table 2) suggested better predictive power for TP compared to CHLa. For

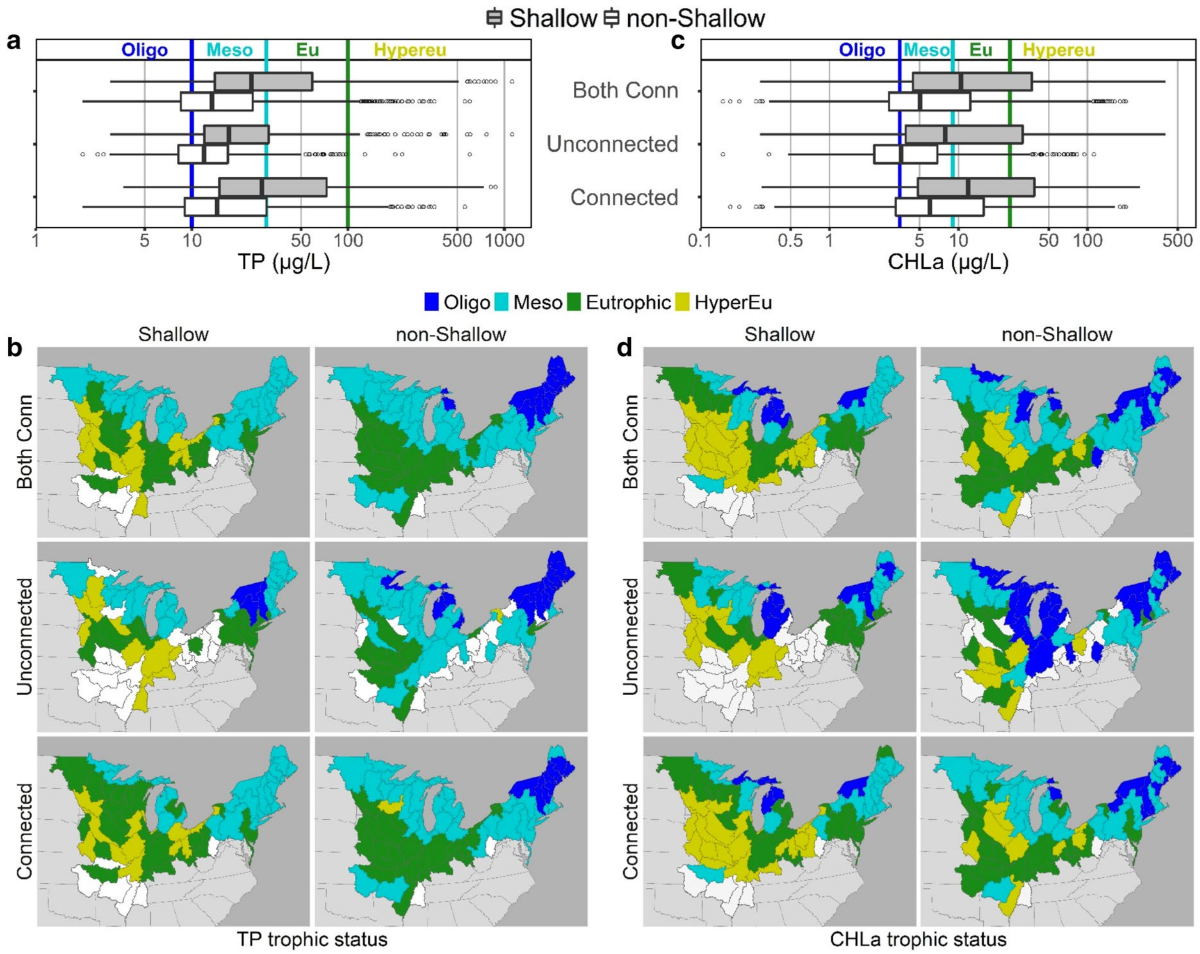


Fig. 4 Boxplots of TP (**A**) and CHLa (**C**) concentration and maps of HU4-specific values of median lake trophic status for TP (**C**) and CHLa (**D**). In the boxplots, the boxes represent 25th and 75th percentiles with the median indicated by a thick black line; whiskers indicate values extending to 1.5 times the interquartile range, and points represent outliers. Vertical colored lines mark transitions in trophic status along each concentration gradient, which are labeled at the top of each boxplot. Concentrations are plotted along a natural logarithmic scale. HU4-specific median trophic status is mapped for TP (**B**) and CHLa (**D**), respectively. Boxplots and maps depict Shallow lakes and non-Shallow lake subsets divided into connectivity classes of connected, unconnected, and both combined (bottom to top rows in both sets of plots). Trophic status categories were defined per Dodds et al. (2006) with cutoffs in µg/l for, respectively, oligotrophic, mesotrophic, eutrophic, and hyper-eutrophic of < 10, 10–30, 30–100, and > 100 for TP and < 3.5, 3.5–9, 9–25, and > 25 for CHLa. White-shaded HU4s do not have any study lakes in that depth by connectivity by trophic variable combination

example, Shallow lake R^2_{cond} values for TP and CHLa were, respectively, 0.60 and 0.50. For both TP and CHLa models, R^2_{cond} values for non-Shallow lakes were higher for connected lakes than for unconnected lakes, but were roughly similar for the two Shallow lake connectivity classes. Regional-scale variation contributed to predictive models as demonstrated by values for R^2_{cond} always exceeding R^2_{marg} (i.e., models with both fixed and random effects were a better fit than those with fixed-effects only); this difference was least pronounced in models of Shallow lake CHLa.

Supporting the patterns in Fig. 4, Shallow lake models of both TP and CHLa had higher intercepts with non-overlapping 95% credible intervals in all paired comparisons with the non-Shallow lake subset, regardless of connectivity class (Table 2, Supplement 2). Intercept estimates were higher for connected than unconnected lakes within both lake depth subsets for

Table 2 Posterior mean parameter estimates, 95% credible intervals (CI), and model fit statistics (root mean square error [rmse], R^2_{marg}, and R^2_{cond}) from individual Bayesian hierarchical linear models for log_e(CHL) and log_e(TP) for Shallow and non-Shallow lakes

Parameters	Shallow lakes			non-Shallow lakes		
	Unconn	Conn	Both Conn	Unconn	Conn	Both Conn
Chlorophyll *a*						
n	614	1330	1944	952	2497	3449
rmse	0.98	0.93	0.95	0.84	0.87	0.87
R^2_{marg} mean	0.48	0.46	0.46	0.24	0.42	0.40
R^2_{marg} 95% CI	(0.43, 0.53)	(0.43, 0.50)	(0.43, 0.49)	(0.20, 0.29)	(0.39, 0.45)	(0.37, 0.42)
R^2_{cond} mean	0.49	0.51	0.50	0.42	0.52	0.49
R^2_{cond} 95% CI	(0.44, 0.55)	(0.47, 0.55)	(0.46, 0.53)	(0.33, 0.52)	(0.48, 0.56)	(0.46, 0.53)
Intercept mean	2.24	2.63	2.54	1.42	1.93	1.79
Intercept 95% CI	(2.03–2.46)	(2.51–2.75)	(2.41–2.66)	(1.18–1.66)	(1.80–2.06)	(1.66–1.93)
Total phosphorus						
n	381	1005	1386	749	2016	2765
rmse	0.68	0.69	0.69	0.55	0.60	0.60
R^2_{marg} mean	0.51	0.57	0.55	0.30	0.56	0.52
R^2_{marg} 95% CI	(0.43, 0.57)	(0.53, 0.60)	(0.52, 0.58)	(0.25, 0.35)	(0.53, 0.59)	(0.49, 0.54)
R^2_{cond} mean	0.64	0.61	0.60	0.42	0.64	0.59
R^2_{cond} 95% CI	(0.56, 0.72)	(0.57, 0.64)	(0.57, 0.64)	(0.32, 0.54)	(0.61, 0.66)	(0.56, 0.62)
Intercept mean	3.35	3.52	3.45	2.57	2.83	2.75
Intercept 95% CI	(3.18–3.52)	3.42–3.61)	(3.36–3.55)	(2.44–2.71)	(2.75–2.92)	(2.68–2.83)

Unconn unconnected lakes, *Conn* connected lakes, *Both conn* combined connectivity classes. Further tests comparing intercept estimates between paired models can be found in Supplement 2

CHLa. In contrast, intercept differences overlapped zero for Shallow lake TP models (Table 2, Supplement 2).

Important predictor variables in hierarchical models of TP and CHLa (i.e., 95% credible intervals of the estimated parameter do not overlap zero) were mainly local scale, with regional-scale predictors rarely important (Fig. 5). For TP, watershed to lake area ratio (+),watershed forest cover (−), and baseflow (−) were important in models of both Shallow and non-Shallow lakes. However, those relationships depended on connectivity class, with fewer variables important in models of unconnected lakes. Lake area (+) and woody wetlands (−) had effects on TP in Shallow connected lakes. Other predictors such as spring (−) or summer (+) precipitation, row crop agriculture (+), and urban land use (+) were occasionally significant in models of non-Shallow lakes, although never for unconnected lakes. Similar to TP, watershed forest cover (−) and baseflow (−) were important predictors of CHLa, regardless of connectivity class or depth subset. In contrast, watershed:lake area ratio, which is associated with nutrient loading, was significant only for predicting non-Shallow lake CHLa. Winter precipitation (−) was important for Shallow and connected lake CHLa models. Unlike TP, regional-scale predictors were occasionally important for CHLa [emergent wetlands (+) and urban land cover (−)], but only in models of unconnected Shallow lakes.

For lakes concurrently sampled for both TP and CHLa, modeled intercepts of the positive relationship between log_e(TP) and log_e(CHLa) were similar regardless of lake connectivity or depth class (Fig. 6; Supplement 3). However, the 95% CI of differences in slope between paired models indicated less steep slopes for Shallow compared to non-Shallow subsets when modeling both connectivity classes combined and connected lakes; 95% CI of the differences overlapped 0 for unconnected lakes (Fig. 6; Supplement 3). All HU4-specific slope estimates (with one exception) from the six models

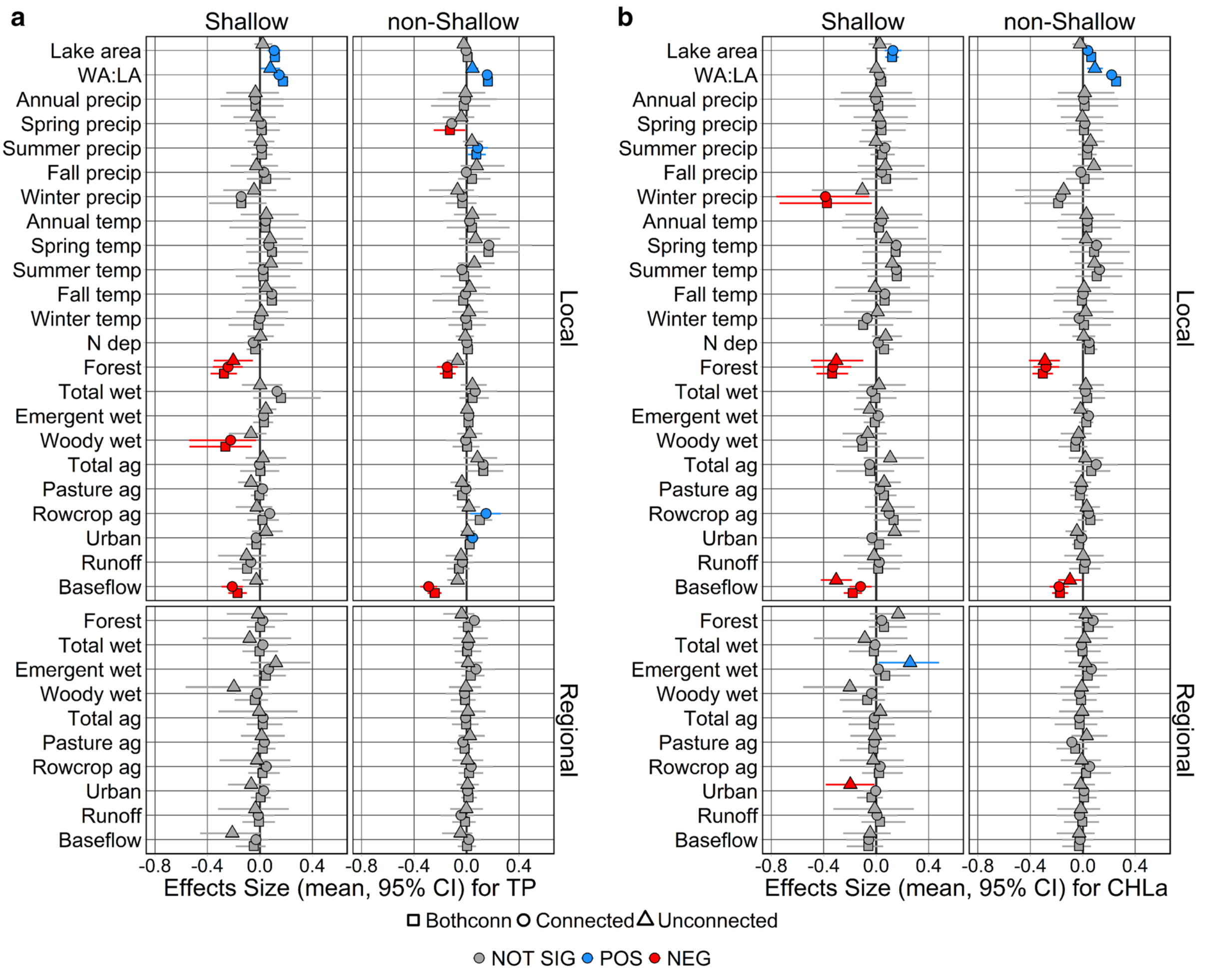


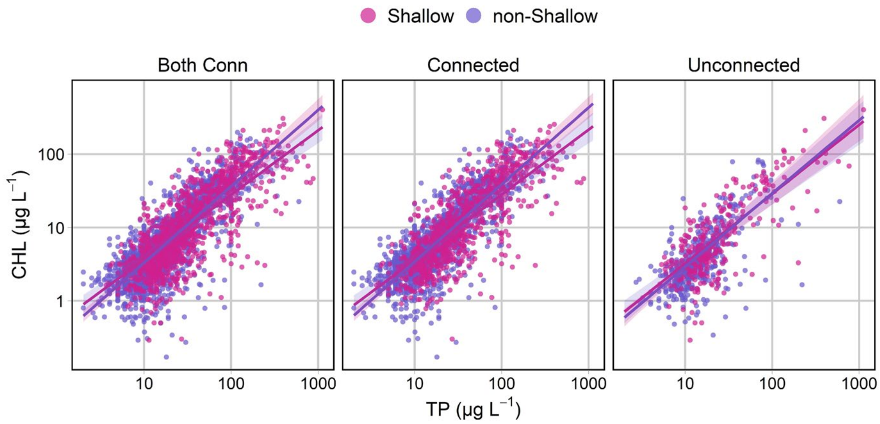
Fig. 5 Effects plots for predictor variables in Bayesian hierarchical models of lake $\log_e$(TP) (left panel) and $\log_e$(CHLa) (right panel) for Shallow and non-Shallow lake subsets. Posterior mean effect sizes and the corresponding 95% credible interval (CI) for predictor variables are plotted for each connectivity group (diamonds are unconnected lakes, circles are connected lakes, and squares are both connectivity classes). Parameters with 95% credible intervals not overlapping zero have either a positive (blue) or negative (red) effect on either TP or CHLa concentration; gray symbols indicate that the credible intervals overlap zero. Predictors are arranged with local or lake-specific scale variables at the top and regional-scale variables at the bottom of each panel. See Supplement 1 for more information about predictor variables

had a 95% CI not overlapping zero; the 95% CI for HU4-specific intercepts did not overlap 0 for 12% to18% and for 25 to 57% of the Shallow and non-Shallow models, respectively (data not shown). This result suggests spatial variation in the HU4-specific TP–CHLa relationship for all models (HU4-specific plots shown in Supplement 4). None of the regional-scale predictors we tested explained any of the among-HU4 variation, thus no cross-scale interactions were detected (Supplement 5).

Discussion

Our macroscale study of Northeastern and Upper Midwestern USA lakes documented differences in trophic status between Shallow and non-Shallow lakes and between connected and unconnected lakes. The results support our expectations that Shallow and connected lakes tend to have higher levels of TP and/or CHLa compared to non-Shallow lakes and unconnected lakes. Further, we found regional variation contributed to predictive models of TP and

Fig. 6 Scatterplots of TP–CHLa relationships plotted on a natural logarithmic scale comparing non-Shallow lakes (purple) and Shallow study lakes (pink) for both connectivity types (left panel), connected lakes (middle panel), and unconnected lakes (right panel). Lines and shading indicate the posterior mean and 95% credible interval for the linear relationships defined by each hierarchical model



CHLa concentrations and that the slope of TP-CHLa relationships varied regionally. Many of the predictor variables that accounted for variation in TP and CHLa among lakes were the same for Shallow and non-Shallow lakes and were consistent with previous macroscale research of lakes across broad spatial gradients. For example, watershed forest cover had a negative effect on TP and CHLa across connectivity classes and depth subset. However, other important predictor variables differed by connectivity or lake depth class. For example, baseflow had a significant negative effect on lake TP in connected but not unconnected lakes, variation among unconnected lakes was not as well explained by our models than in connected lake models, and the regional-scale variables were unimportant with the exception of emergent wetland and urban land use in models of CHLa for unconnected Shallow lakes.

Across all models, TP–CHLa relationships were positive and were not affected by cross-scale interactions. Others found that non-linear models were more appropriate across wide ranges of TP and CHLa (Filstrup et al., 2014; Quinlan et al., 2021), with CHLa potentially limited by nitrogen at very low TP (e.g., in alpine settings) and co-limited by nitrogen at very high TP concentrations (Filstrup & Downing, 2017; Quinlan et al., 2021). However, ultra-oligotrophic lakes were not present in our study lakes. In fact, the majority of lakes were within the intermediate range of 4–230 ug/l TP, within which the TP–CHLa relationship is relatively linear (Quinlan et al., 2021). Although this fact supports our use of a linear hierarchical modeling approach, our finding that shallow lakes are more commonly hyper-eutrophic suggests that non-linear models of TP–CHLa are likely more appropriate in future studies across a wide range of trophic status. Our linear TP–CHLa models had steeper slopes for non-Shallow lakes than Shallow lakes, which is opposite that found by Quinlan et al. (2021). This discrepancy could be an artifact of our linear model being more influenced by relatively higher TP and CHLa in Shallow than non-Shallow lakes, a result not observed in the asymptotic models used by Quinlan et al. (2021).

The distributions of TP and CHLa concentration across our study extent were characterized by increasing concentrations and trophic status from north to south, very similar to that observed at the continental US scale for Secchi transparency (Bigham Stephens et al., 2015). Lapierre et al. (2018) quantified the spatial structure of lake and ecological setting variables found weak spatial structure for lake-specific variables such as depth, intermediate spatial structuring of TP and CHLa, and strong spatial structuring of ecological setting variables characterizing climate, runoff, and land use/cover. Despite finding similarly strong spatial patterns in TP and CHLa, we detected no relationships between region-specific slopes and regional-scale predictors of land use/cover and baseflow. This result contrasts with other studies that suggest the yield of CHLa per unit TP differs according to both local and regional settings. For example, studies have detected relationships between region-specific slopes of the TP–CHLa relationship and regional predictor variables such as percent pasture (Wagner et al., 2011) and wetland cover (Filstrup et al., 2014).

Our results support other studies that found lake trophic states and relationships were related to climate variables, which are spatially structured. For example, recent studies of global patterns of the TP–CHLa relationship (Quinlan et al., 2021) and on CHLa spatial patterns (Shuvo et al., 2021) suggest that not only do deeper lakes have lower CHLa per unit TP than predicted, but also that complex north–south patterns related to climate produce higher CHLa in extremely warm climates. Although our study extent is considered north temperate, climate-related variables such as winter precipitation (CHLa) and baseflow (both TP and CHLa) were important local drivers of concentration. Research also suggests potential for increasing CHLa production under climate warming scenarios (Collins et al., 2017; Quinlan et al., 2021; and Shuvo et al., 2021). Because shallow lakes are influenced by complex interactions between macrophytes, foodweb structure, and humic substances, all of which may be related to large-scale climate gradients (Kosten et al., 2009), it follows that responses to climate change may be particularly critical for shallow lakes.

Differences in model results between Shallow and non-Shallow lakes reinforce the importance of accounting for lake depth in future macroscale studies of lake states and relationships. Only 23% of lakes within our study extent had depth data, and although 40% of those lakes were shallow, only 34% of those lakes had measurements of either TP or CHLa (with only 26% having *both* TP and CHLa). Efforts to quantify the biases present in compiled datasets such as LAGOS-NE have found that larger lakes are more likely to be sampled (Wagner et al., 2008; Stanley et al., 2019). The bias against sampling shallow lakes may be even larger for the entire population of lakes within our study extent since the smallest lakes, which are thought to be disproportionately shallow, are not included in LAGOS-NE (the minimum lake area equal to 4 hectares; Soranno et al., 2017). In fact, the 2012 US National Lakes Assessment that includes lakes > 1 ha in surface area estimates that 61% of US lakes may be shallow (USEPA, 2016). In addition to the bias toward deeper lakes, depth data were more available for connected than for unconnected lakes in the LAGOS-NE study extent, which is biased toward both larger and connected lakes (Stanley et al., 2019). Therefore, our understanding of lake TP and CHLa in this lake-dense area of the world is affected by the disproportionate under-sampling of shallow, unconnected lakes. Given that small lakes, which are often shallow, are important for carbon cycling and other global processes (Downing, 2010; Holgerson & Raymond, 2016; Biggs et al., 2017), the need to increase our understanding of the controls on nutrient status of these lakes is even more pressing.

Taking a macroscale approach to study thousands of lakes in the Northeastern and Upper Midwestern USA allowed us to demonstrate the generality of shallow lakes having higher TP and CHLa than deeper lakes, as well as nuances associated with connectivity and region. There are many newly compiled and created macroscale datasets quantifying connectivity and regional predictor variables important for modeling TP and CHLa and their relationships (e.g., Hill et al., 2018; Cheruvelil et al., 2021; King et al., 2021). However, we lack lake depth data for most ecosystems in the USA (Stachelek et al., 2022; Webster et al. 2022) and lake depth cannot be remotely sensed, is time intensive to generate, and is notoriously difficult to estimate using predictive models (e.g., Hollister et al., 2011; Sobek et al., 2011; Oliver et al., 2016). This general paucity of lake depth data has important implications. Estimates of global lake contributions to carbon and water cycles based on only large (and mostly deep) lakes are missing an important population of lakes (Downing, 2010; Holgerson & Raymond, 2016; Biggs et al., 2017). Because shallow lakes experience frequent turnover and have tightly linked benthic-pelagic processes, TP and CHLa concentrations may be more sensitive to and respond differently and at different rates to land use intensification and climate change than deeper lakes. Therefore, this lack of depth data may be affecting limnologists' understanding of basic patterns and drivers, as well as responses to global changes. If better understanding the response of shallow lakes to global change and their contribution to global cycles is a research objective, then future research prioritizing lake depth as a variable to quantify and the study of shallow and unconnected lakes may be warranted.

Acknowledgements Thanks to the members of the CSI Limnology and Continental Limnology research groups, the organizing committees and participants of the 2014 and 2021 Shallow Lakes Congresses, Patricia Soranno, and the anonymous reviewers. Funding was provided by the US National Science Foundation (2011–2017: EF-1065786; 2016-present: DEB-1638679 and DEB-1638539). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

Author contributions The initial conception for this manuscript came from KSC's preparation for the 2014 Shallow Lakes Congress in Antalya, Turkey. KSC and KEW are co-lead authors and led the team in writing. KSC, KEW, and TW led the team in analysis and model interpretation, with TW writing the statistical modeling methods. Using earlier versions of the data, KBSK and ACP conducted data processing, data analysis, made tables and figures, and assisted with literature review under KSC's mentorship. KEW conducted final model runs and made the final figures and tables. All co-authors reviewed and revised the entire manuscript.

Funding Funding was provided by the US National Science Foundation (2011–2017: EF-1065786; 2016-present: DEB-1638679 and DEB-1638539).

Data availability Cheruvelil, K.S., K.E. Webster, K.B. King, A.C. Poisson, and T. Wagner. 2022. LAGOS-NE Shallow Lakes: a dataset of lake variables and multi-scaled ecological context variables used to predict and compare trophic status and TP:CHLa relationships between shallow and non-shallow lakes in the Upper Midwest and Northeastern United States. ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/be49507b941815d7a6807a273ee02d1e (Accessed 2022-02-09).

Code availability Available at the same location as the data.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

Consent to participate N/a.

Ethical approval N/a.

References

Bigham Stephens, D. L., R. E. Carlson, C. A. Horsburgh, M. V. Hoyer, R. W. Bachmann & D. E. Canfield Jr., 2015. Regional distribution of Secchi disk transparency in waters of the United States. Lake and Reservoir Management 31(1): 55–63.

Brett, M. T. & M. M. Benjamin, 2008. A review and reassessment of lake phosphorus retention and the nutrient loading concept. Freshwater Biology 53: 194–211.

Biggs, J., S. von Fumetti & M. Kelly-Quinn, 2017. The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. Hydrobiologia 793: 3–39.

Brooks, S. P. & A. Gelman, 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7: 434–455.

Cheruvelil, K. S. & P. A. Soranno, 2018. Data-intensive ecological research is catalyzed by open science and team science. BioScience 68(19): 813–822.

Cheruvelil, K. S., K. E. Webster, K. B. King, A. C. Poisson, & T. Wagner. 2022. LAGOS-NE Shallow Lakes: a dataset of lake variables and multi-scaled ecological context variables used to predict and compare trophic status and TP:CHLa relationships between shallow and non-shallow lakes in the Upper Midwest and Northeastern United States. ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/be49507b941815d7a6807a273ee02d1e. Accessed 09 02 2022.

Collins, S. M., S. K. Oliver, J. Lapierre, E. H. Stanley, J. R. Jones, T. Wagner & P. A. Soranno, 2017. Lake nutrient stoichiometry is less predictable than nutrient concentrations at regional and sub-continental scales. Ecological Applications 27: 1529–1540.

Collins, S. M., S. Yuan, P. N. Tan, S. K. Oliver, J. F. Lapierre, K. S. Cheruvelil, C. E. Fergus, N. K. Skaff, J. Stachelek, T. Wagner & P. A. Soranno, 2019. Winter precipitation and summer temperature predict lake water quality at macroscales. Water Resources Research 55: 2708–2721. https://doi.org/10.1029/2018WR023088.

Cheruvelil, K. S., P. A. Soranno, I. M. McCullough, K. E. Webster, L. K. Rodriguez & 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. Limnology and Oceanography Letters 6: 270–292.

Dodds, W. K., E. Carney & R. T. Angelo, 2006. Determining ecoregional reference conditions for nutrients, Secchi depth and chlorophyll *a* in Kansas lakes and reservoirs. Lake and Reservoir Management 22(2): 151–159.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitao & T. Münkemüller, 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27–46.

Downing, J. A., 2010. Emerging global role of small lakes and ponds: little things mean a lot. Limnetica 29(1): 9–24.

Filstrup, C. T. & J. A. Downing, 2017. Relationship of chlorophyll to phosphorus and nitrogen in nutrient-rich lakes. Inland Waters 7(4): 385–400.

Filstrup, C. T., T. Wagner, P. A. Soranno, E. H. Stanley, C. A. Stow, K. E. Webster & J. A. Downing, 2014. Regional variability among nonlinear chlorophyll-phosphorus relationships in lakes. Limnology and Oceanography 59: 1691–1703.

Gelman, A. & J. Hill, 2007. Data analysis using regression and multilevel/hierarchical models, Cambridge University Press, New York:

Heffernan, J. B., P. A. Soranno, M. J. Angilletta, L. B. Buckley, D. S. Gruner, T. H. Keitt, J. R. Kellner, J. S. Kominoski, A. V. Rocha, J. Xiao, T. K. Harms, S. J. Goring, L. E. Koenig, W. H. McDowell, H. Powell, A. D. Richardson, C. A. Stow, R. Vargas & K. C. Weathers, 2014. Macrosystems ecology: understanding ecological patterns and processes at continental scales. Frontiers in Ecology and the Environment 12: 5–14.

Hill, R. A., M. H. Weber, R. M. Debbout, S. G. Leibowitz & A. R. Olsen, 2018. The Lake-Catchment (LakeCat) dataset: characterizing landscape features for lake basins within the conterminous USA. Freshwater Science 37(2): 208–221.

Holgerson, M. & P. Raymond, 2016. Large contribution to inland water CO_2 and CH_4 emissions from very small ponds. Nature Geoscience 9: 222–226. https://www.nature.com/articles/ngeo2654.

Hollister, J. W., W. B. Milstead & M. A. Urrutia, 2011. Predicting maximum lake depth from surrounding topography. PLOS ONE Public Library of Science 6: e25764.

Jackson, L. J., T. L. Lauridsen, M. Søndergaard & E. Jeppesen, 2007. A comparison of shallow Danish and Canadian lakes and implications of climate change. Freshwater Biology 52: 1782–1792.

Jeppesen, E., J. P. Jensen, M. Søndergaard, T. Lauridsen, L. J. Pedersen & L. Jensen, 1997. Top-down control in freshwater lakes: the role of nutrient state, submerged macrophytes and water depth. Hydrobiologia 342: 151–164.

Kalff, J., 2001. Limnology: inland water ecosystems, Prentice Hall, Hoboken:

Kellner, K., 2019. jagsUI: a wrapper around 'rjags' to streamline 'JAGS' analyses. R package version 1.5.1. https://CRAN.R-project.org/package=jagsUI

King, K. B. S., Q. Wang, L. K. Rodriguez & K. S. Cheruvelil, 2021. Lake networks and connectivity metrics for the conterminous U.S. (LAGOS-US NETWORKS v1). Limnology and Oceanography Letters 6: 293–307.

Kosten, S., G. Lacerot, E. Jeppesen, D. da Motta Marques, E. van Nes, N. Mazzeo & M. Scheffer, 2009. Effects of submerged vegetation on water clarity across climates. Ecosystems 12(7): 1117–1129.

Lapierre, J.-F., S. M. Collins, D. A. Seekell, K. S. Cheruvelil, P.-N. Tan, N. K. Skaff, Z. Taranu, C. E. Fergus & P. A. Soranno, 2018. Similarity in spatial structure constrains ecosystem relationships: building a macroscale understanding of lakes. Global Ecology and Biogeography 27(10): 1251–1263.

Martin, S. L. & P. A. Soranno, 2006. Lake landscape position: Relationships to hydrologic connectivity and landscape features. Limnology and Oceanography 51: 801–814.

Nakagawa, S. & H. Schielzeth, 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4: 133–142.

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

Padisák, J. & C. S. Reynolds, 2003. Shallow lakes: the absolute, the relative, the functional and the pragmatic. Hydrobiologia 506–509: 1–11.

Park, T. & G. Casella, 2008. The Bayesian Lasso. Journal of the American Statistical Association 103: 681–686.

Phillips, G., O.-P. Pietiläinen, L. Carvalho, A. Solimini, A. Lyche Solheim & A. C. Cardoso, 2008. Chlorophyll–nutrient relationships of different lake types using a large European dataset. Aquatic Ecology 42: 213–226.

Plummer M., 2003. JAGS: A program for analysis of bayesian graphical models using gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (eds), Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003). Vienna, Austria. ISSN 1609-395X. http://www.ci.tuwien.ac.at/Conferences/DSC-2003/.

Quinlan, R., A. Filazzola, O. Mahdiyan, A. Shuvo, K. Blagrave, C. Ewins, L. Moslenko, D. K. Gray, C. M. O'Reilly & S. Sharma, 2021. Relationships of total phosphorus and chlorophyll in lakes worldwide. Limnology and Oceanography 66: 392–404.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Read, E. K., V. P. Patil, S. K. Oliver, A. L. Hetherington, J. A. Brentrup, J. A. Zwart, K. M. Winters, J. R. Corman, E. R. Nodine & R. I. Woolway, 2015. The importance of lake-specific characteristics for water quality across the continental United States. Ecological Applications 25: 943–955.

Scheffer, M., 1998. Ecology of Shallow Lakes, Springer, Berlin, Germany.

Scheffer, M., S. H. Hosper, M. L. Meijer, B. Moss & E. Jeppesen, 1993. Alternative equilibria in shallow lakes. Trends in Ecology and Evolution 8: 275–279.

Seaber, P. R., F. P. Kapinos & G. L. Knapp, 1987. Hydrologic unit map, United States Geological Survey, Reston, VA, USA.

Shuvo, A., C. M. O'Reilly, K. Blagrave, C. Ewins, A. Filazzola, D. Gray, O. Mahdiyan, L. Moslenko, R. Quinlan & S. Sharma, 2021. Total phosphorus and climate are equally important predictors of water quality in lakes. Aquatic Sciences 83(1): 1–11.

Sobek, S., J. Nisell & J. Fölster, 2011. Predicting the depth and volume of lakes from map-derived parameters. Inland Waters 1: 177–184.

Soranno, P. A., K. E. Webster, J. L. Riera, T. K. Kratz, J. S. Baron, P. A. Bukaveckas, G. W. Kling, D. S. White, N. Caine, R. C. Lathrop & P. R. Leavitt, 1999. Spatial variation among lakes within landscapes: ecological organization along lake chains. Ecosystems 2: 395–410.

Soranno, P. A., K. S. Cheruvelil, E. G. Bissell, M. T. Bremigan, J. A. Downing, C. E. Fergus, C. T. Filstrup, E. N. Henry, N. R. Lottig, E. H. Stanley, C. A. Stow, P.-N. Tan, T. Wagner & K. E. Webster, 2014. Cross-scale interactions: quantifying multi-scaled cause–effect relationships in macrosystems. Frontiers in Ecology and the Environment 12: 65–73.

Soranno, P. A., E. G. Bissell, K. S. Cheruvelil, S. T. Christel, S. M. Collins, C. E. Fergus, C. T. Filstrup, J.-F. Lapierre, N. R. Lottig, S. K. Oliver, C. E. Scott, N. J. Smith, S. Stopyak, S. Yuan, M. T. Bremigan, J. A. Downing, C. Gries, E. N. Henry, N. K. Skaff, E. H. Stanley, C. A. Stow, P.-N. Tan, T. Wagner & K. E. Webster, 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse. GigaScience 4: 28.

Soranno, P. A., L. C. Bacon, M. Beauchene, K. E. Bednar, E. G. Bissell, C. K. Boudreau, M. G. Boyer, M. T. Bremigan, S. R. Carpenter, J. W. Carr, K. S. Cheruvelil, S. T. Christel, M. Claucherty, S. M. Collins, J. D. Conroy, J. A. Downing, J. Dukett, C. E. Fergus, C. T. Filstrup, C. Funk, M. J. Gonzalez, L. T. Green, C. Gries, J. D. Halfman, S. K. Hamilton, P. C. Hanson, E. N. Henry, E. M. Herron, C. Hockings, J. R. Jackson, K. Jacobson-Hedin, L. L. Janus, W. W. Jones, J. R. Jones, C. M. Keson, K. B. S. King, S. A. Kishbaugh, J.-F. Lapierre, B. Lathrop, J. A. Latimore,

Y. Lee, N. R. Lottig, J. A. Lynch, L. J. Matthews, W. H. McDowell, K. E. B. Moore, B. P. Neff, S. J. Nelson, S. K. Oliver, M. L. Pace, D. C. Pierson, A. C. Poisson, A. I. Pollard, D. M. Post, P. O. Reyes, D. O. Rosenberry, K. M. Roy, L. G. Rudstam, O. Sarnelle, N. J. Schuldt, C. E. Scott, N. K. Skaff, N. J. Smith, N. R. Spinelli, J. J. Stachelek, E. H. Stanley, J. L. Stoddard, S. B. Stopyak, C. A. Stow, J. M. Tallant, P.-N. Tan, A. P. Thorpe, M. J. Vanni, T. Wagner, G. Watkins, K. C. Weathers, K. E. Webster, J. D. White, M. K. Wilmes & S. Yuan, 2017. LAGOS-NE: a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes. GigaScience 6: 1–22.

Stachelek, J. J., W. Weng, C. C. Carey, A. R. Kemanian, K. M. Cobourn, T. Wagner, K. C. Weathers & P. A. Soranno, 2020. Granular measures of agricultural land use influence lake nitrogen and phosphorus differently at macroscales. Ecological Applications 30(8): e02187.

Stachelek, J., L. K. Rodriguez, J. Díaz Vázquez, A. Hawkins, E. Phillips, A. Shoffner, I. M. McCullough, K. B. King, J. Namovich, L. A. Egedy, M. Haite, P. J. Hanly, K. E. Webster, K. S. Cheruvelil & P. A. Soranno, 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. Limnology and Oceanography Letters 6: 270–292.

Stachelek, J., P. J. Hanly & P. A. Soranno, 2022. Imperfect slope measurements drive overestimation in geometric cone model of lake and reservoir depth. Inland Waters. https://doi.org/10.1080/20442041.2021.2006553.

Stanley, E. H., S. M. Collins, N. R. Lottig, S. K. Oliver, K. E. Webster, K. S. Cheruvelil & P. A. Soranno, 2019. Biases in lake water quality sampling and implications for macroscale research. Limnology and Oceanography 64: 1572–1585.

Taranu, Z. E. & I. Gregory-Eaves, 2008. Quantifying relationships among phosphorus, agriculture, and lake depth at an inter-regional scale. Ecosystems 11: 715–725.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B 58: 267–288.

USEPA, 2016. National lakes assessment 2012: a collaborative survey of lakes in the United States. EPA 841-R-16–113. U.S. Environmental Protection Agency, Washington, D.C., USA.

Vander Zanden, M. J. & Y. Vadeboncoeur, 2020. Putting the lake back together 20 years later: what in the benthos have we learned about habitat linkages in lakes? Inland Waters 10(3): 305–321.

Vollenweider, R. A., 1975. Input-output models with special reference to the phosphorus loading concept in limnology. Schweizerische Zeitschrift Für Hydrologie 37: 53–84.

Wagner, T., P. A. Soranno, K. S. Cheruvelil, W. H. Renwick, K. E. Webster, P. Vaux & R. J. F. Abbitt, 2008. Quantifying sample biases of inland lake sampling programs in relation to lake surface area and land use/cover. Environmental Monitoring and Assessment 141: 131–147.

Wagner, T., P. A. Soranno, K. E. Webster & K. S. Cheruvelil, 2011. Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes: relationship between total phosphorus and chlorophyll. Freshwater Biology 56: 1811–1824.

Wagner, T., C. E. Fergus, C. A. Stow, K. S. Cheruvelil & P. A. Soranno, 2016. The statistical power to detect cross-scale interactions at macroscales. Ecosphere 7: e01417.

Webster, K. E., I. M. McCullough & P. A. Soranno, 2022. Deeper by the Dozen: Diving into a Database of 17,675 Depths for U.S. Lakes and Reservoirs. Limnology and Oceanography Bulletin. https://doi.org/10.1002/lob.10482.

Zhang, T., P. A. Soranno, K. S. Cheruvelil, D. B. Kramer, M. T. Bremigan & A. Ligmann-Zielinska, 2012. Evaluating the effects of upstream lakes and wetlands on lake phosphorus concentrations using a spatially-explicit model. Landscape Ecology 27: 1015–1030.