ADVANCING EARTH AND SPACE SCIENCES

Earth's Future

RESEARCH ARTICLE
10.1029/2024EF004860

Key Points:

- China's lake nutrient concentrations can be accurately predicted by watershed attributes
- China's lake nutrient levels depended on watershed features within a <45 km hydrological distance
- Human footprint increases the relevant hydrological distance to predict water quality for China's lakes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Huang and G. B. Arhonditsis,
jchuang@niglas.ac.cn;
george.arhonditsis@utoronto.ca

Citation:

Huang, J., Arhonditsis, G. B., Zhang, Y., Zhang, S., Ji, Y., Paerl, H. W., et al. (2024). The critical role of hydrological distance in shaping nutrient dynamics along the watershed-lake continuum. *Earth's Future*, *12*, e2024EF004860. https://doi.org/10.1029/2024EF004860

Received 6 MAY 2024
Accepted 4 SEP 2024

The Critical Role of Hydrological Distance in Shaping Nutrient Dynamics Along the Watershed-Lake Continuum

Jiacong Huang[1,2], George B. Arhonditsis[3], Yinjun Zhang[4], Shuai Zhang[1,5], Yulai Ji[1,5], Hans W. Paerl[6], Erik Jeppesen[7,8,9,10], and Junfeng Gao[1]

[1]Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, [2]Poyang Lake Wetland Research Station, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Jiujiang, China, [3]Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, ON, Canada, [4]China National Environmental Monitoring Centre, Beijing, China, [5]University of Chinese Academy of Sciences, Beijing, China, [6]Institute of Marine Sciences, University of North Carolina at Chapel Hill, Morehead City, NC, USA, [7]Department of Bioscience, Aarhus University, Aarhus, Denmark, [8]Sino-Danish Centre for Education and Research, Beijing, China, [9]Limnology Laboratory, Department of Biological Sciences and Centre for Ecosystem Research and Implementation, Middle East Technical University, Ankara, Turkey, [10]Institute of Marine Sciences, Middle East Technical University, Mersin, Turkey

Abstract Terrestrial hydrological and nutrient cycles are subjected to major disturbances by agricultural operations and urbanization that profoundly influence freshwater resources. Non-point source pollution is one of the primary causes for water quality deterioration, and thus an emerging imperative in limnology is establishing empirical models that connect watershed attributes and hydrological drivers with lake nutrient dynamics. Here, we compiled three nation-wide nutrient, meteorological, and watershed-landscape data sets, to develop Generalized Linear Models that predict lake phosphorus and nitrogen concentrations as a function of the surrounding watershed characteristics within various hydrological distances across 104 Chinese lakes and reservoirs. Our national-scale investigation revealed that lake nutrient concentrations can be satisfactorily predicted by proxies of natural drivers and anthropogenic activities, reflecting the properties of the surrounding watershed. Counter to previous studies, we found that China's lake nutrient concentrations strongly depend on watershed characteristics within a hydrological distance of less than 45 km rather than the entire watershed. Furthermore, extensive human activities in watersheds not only compromise our predictive capacity, but also increase the hydrological distance that is relevant to predict lake nutrients. This national-scale characterization can inform one of the most contentious issues in the context of China's lake management, that is, the determination of the extent of the nearshore area, where nutrient control should be prioritized. As far as we know, our study represents the first attempt to apply the concept of hydrological distance and establish statistical models that can delineate the critical spatial domain primarily responsible for the nutrient conditions along the watershed-lake continuum.

Plain Language Summary Connecting watershed attributes with lake water quality at multiple scales is an emerging imperative in limnology. This study introduces the concept of hydrological distance and establish lake-watershed relationships based on a comprehensive water quality data set from 2,424 sampling sites across China. Our key finding is that lake nutrient levels strongly depend on watershed features within a <45 km hydrological distance rather than the entire watershed. Activities related to agriculture and urbanization increase the relevant hydrological distance to modulate lake nutrient levels, and our national-scale analysis provides evidence that areas with discernible impact can be located far from the vicinity of a given lake.

1. Introduction

Anthropogenic activities (e.g., urbanization processes and agricultural operations) profoundly alter the hydrological and nutrient cycles of terrestrial systems, and are intricately linked to both availability and distribution of freshwater resources (Arnillas et al., 2021). Point sources pollution have historically been a primary cause of freshwater degradation. However, during the past decades, wastewater treatment plants have effectively mitigated their impacts. Therefore, nonpoint sources pollution has instead become the leading cause for the surface water quality deterioration. Excess phosphorus (P) and nitrogen (N) loading from watersheds to lakes via rivers is a continuing problem due to its linkage to a wide range of water quality problems including hypoxia, phytoplankton blooms, biodiversity loss and fisheries degradation (Ho et al., 2019; Ibáñez & Peñuelas, 2019; Jane et al., 2021).

10.1029/2024EF004860

Existing global estimates signify the delivery of large amounts of P (4–11 Tg/yr) and N (37–66 Tg/yr) from land into receiving watercourses, with agriculture (>50%) being the dominant contributor (Beusen et al., 2016, 2022; Sharples et al., 2017). Such high nutrient loading are responsible for a discernible increase in the magnitude and frequency of phytoplankton blooms in 68% of lakes since the 1980s (Ho et al., 2019). Under the context of climate change, N loading is estimated to increase 19% ± 14% within the United States during the 21st century, driven by the projected increase in the frequency of precipitation extremes (Sinha et al., 2017).

Excess nutrient load is particularly severe in China, partly due to the excessive use of fertilizers (80 kg P ha^{-1} yr^{-1} and 305 kg N ha^{-1} yr^{-1}) in millions of small farms compared to the worldwide average of fertilizer use (12 kg P ha^{-1} yr^{-1} and 74 kg N ha^{-1} yr^{-1}), with >50% leaching into receiving waters or emitting into atmosphere (Cui et al., 2018; Liu et al., 2016; Lu & Tian, 2017; Zhang et al., 2013). To mitigate the severity of eutrophication problems, water quality managers have made substantial and diverse efforts, for example, construction/restoration of wetlands and wastewater treatment plants (WWTPs), with particular emphasis on three large and highly eutrophic lakes (i.e., Lakes Dianchi, Taihu and Chaohu) and their surrounding watersheds (Huang et al., 2018; Qin et al., 2019; Wu et al., 2017). Several laws and guidelines (e.g., Water Pollution Control Action plan released on 16 April. 2015) delineated the framework for N and P control over the past decades. This legislative framework provided the roadmap for considerable restoration efforts, such as the increasing wastewater treatment plants from 1,535 units in 2008 to 4,960 units in 2017 (Tong et al., 2020). Through these remedial actions, harmful algal blooms have been reduced, but not fully eliminated (Huang et al., 2019). They remain an on-going challenge for China's lakes with an areal magnitude as high as 775.4 km^2 (Huang, Zhang, et al., 2020).

The severity of nonpoint-source pollution is modulated by a complex interplay among anthropogenic activities, landscape attributes and transport pathways, and may be further amplified by climatic change (Huisman et al., 2018; Hundey et al., 2016). Unlike point sources though, the diffuse and intermittent nature of pollution from agricultural areas and urban settings makes it more difficult to monitor, quantify, and even more so to integrate watershed and aquatic processes (Yin et al., 2022; Yuan et al., 2021). Environmental modelling has been one of the pillars of the integrated watershed management, serving as an "information integrator" that brings managers, scientists, and other stakeholders together to answer our knowledge gaps (Huang et al., 2018). Previous studies attempted to link lake water quality with watershed characteristics via data-driven and mechanistic modeling (Huang et al., 2018; Peng et al., 2021; Wei et al., 2020). Data-driven models (e.g., Generalized Linear Models (GLMs) and Bayesian hierarchical models) attempted to link water quality of a lake/river with its corresponding watershed or surrounding area defined by a buffering distance (Huang et al., 2021). These empirical models are simple linear (or non-linear) relationships typically based on predictor variables (e.g., catchment attributes, landscape characteristics, meteorological forcing) that are readily available, and are thus perfectly tailored for broad-scale investigations across multiple watersheds, where detailed data may not always be available. Nonetheless, the simple nature of empirical models is also the primary reason for their limited use in more comprehensive granular investigations, especially within individual catchments, as they fail to capture the complexity of the processes underlying nutrient cycling during the transport from watersheds to lakes via rivers, streams, and ditches. To address the structural and conceptual weaknesses of statistical models, another option is to use process-based (mechanistic) modeling to describe surface runoff, biogeochemical cycling, sediment transport, and river routing along the watershed-lake continuum (Arhonditsis et al., 2019a; Janssen et al., 2019). Process-based models require substantial input data, and are generally used for site-specific applications rather than broad-scale studies and national-scale assessments (Huang et al., 2018). Thus, notwithstanding our evolving understanding of the key facets of non-point source pollution, there is a multitude of "management-relevant" questions related to the delineation of high-risk areas in anthropogenically disturbed watersheds and the identification of locations associated with excessive nutrient load generation and high delivery rates, as well as periods of the year with high levels of risk (or uncertainty) for terrestrial and aquatic ecosystem resilience.

This study aims to establish linkages between a suite of natural and anthropogenic watershed attributes and in-lake N and P levels. A hydrological-distance analysis framework was developed to identify the locations, where the signature of the watershed landscape features on lake nutrient conditions is mostly discernible. For this purpose, we compiled three comprehensive lake nutrient concentration, meteorological, and watershed-landscape data sets, to develop GLMs that predict lake N and P conditions as a function of the surrounding watershed characteristics within various hydrological distances across 104 Chinese lakes and reservoirs. Empirical models were then developed to estimate nutrient attenuation rates from watersheds to lakes. To the best of our knowledge, the present study represents the first attempt to use the concept of hydrological distance in an applied context in

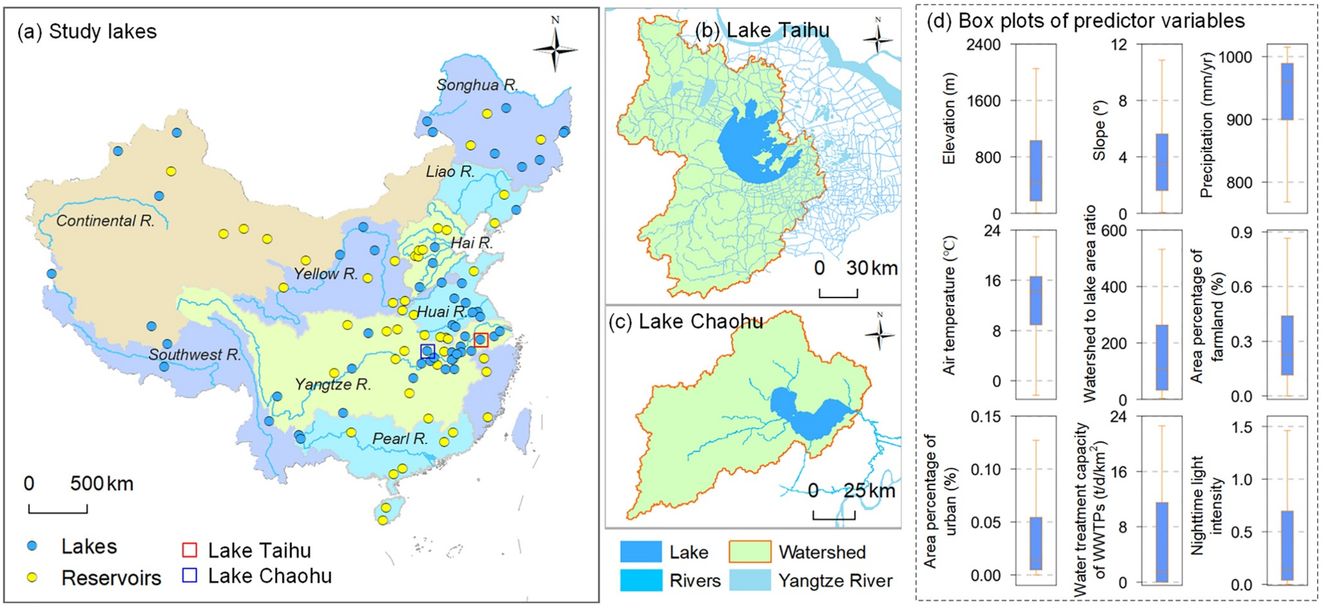

Figure 1. (a) Nationwide monitoring network in China's lakes/reservoirs with two examples of lake-watershed systems, that is, (b) Lakes Taihu and (c) Chaohu, and (d) the distribution of the values of nine predictor variables for 104 lake-watershed ecosystems across China.

order to identify not only the most sensitive predictors, but also to delineate the critical spatial domain that shapes nutrient dynamics along the watershed-lake continuum.

2. Materials and Methods

2.1. Study Area and Data

In China, there are 2,693 lakes with a surface area of >1 km^2, with an increasing number of lakes during the past decades (Ma et al., 2011; Zhang et al., 2019). Among these lakes, ∼142 lakes (including reservoirs) across China are of concern because they are heavily impacted by human activities (Huang et al., 2019). To investigate the relationships between lake N and P levels and watershed characteristics, the following three nationwide data sets were compiled.

2.1.1. Lake Nutrient Data Set

To capture lake nutrient conditions across the nation, Chinese National Environmental Monitoring Center (http://www.cnemc.cn/) conducted a monthly sampling program in China's 142 lakes (including reservoirs). Among these, 104 lakes (including 48 reservoirs) have a watershed with hydrological distance >50 km, and were selected as the study sites (Figure 1). Based on the sampling program, a large data set was collected including numerous monthly samples during the 2016–2020 period. This data set included 24 water quality variables listed as the environmental quality standards for surface water (GB 3838–2002). This study shed lights on total nitrogen (TN), total ammonia nitrogen (NH_3–N, the sum of free ammonia and ammonium), and total phosphorus (TP).

2.1.2. Meteorological Data Set

To capture the role of meteorological conditions in shaping lake nutrient concentrations, the Chinese Meteorological Forcing Data set (CMFD) with a 0.1° spatial and daily temporal resolution was used. The data set with gridded near-surface meteorological data was specifically developed for studies of land-surface processes (https://data.tpdc.ac.cn/). The data set includes seven variables: precipitation (Pr, mm/d), pressure (Pa), air temperature (T, °C), specific humidity (kg/kg), wind speed (m/s), downward shortwave and downward longwave radiation (W/m^2) (He et al., 2020).

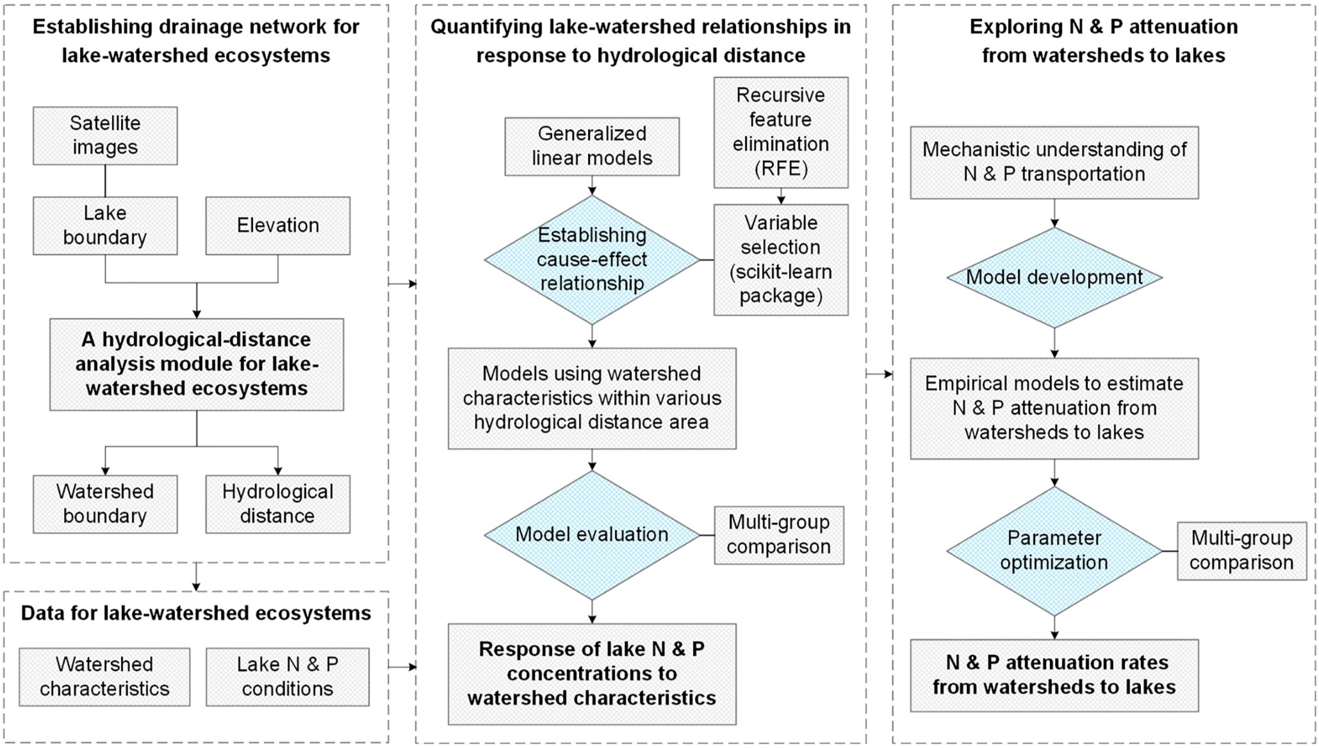

Figure 2. Schematic illustration of our modeling analysis on the response of lake phosphorus (P) and nitrogen (N) concentrations to watershed characteristics.

2.1.3. Watershed Landscape Data Set

The data set includes eight variables: land use, elevation, slope, precipitation, air temperature, water treatment capacity of WWTPs, nighttime light intensity, and watershed-to-lake ratio (Figures S1 and S2 in Supporting Information S1). Land use was obtained from the Resource and Environment Data Cloud Platform (http://www.resdc.cn/Default.aspx). Elevation data were obtained from Shuttle Radar Topography Mission (SRTM, http://srtm.csi.cgiar.org/srtmdata/). Slope data were derived using the elevation data. Meteorological data (precipitation and air temperature) were obtained from CMFD. Nighttime light intensity was obtained from the Defense Meteorological Satellite Program (DMSP) satellites (https://www.ngdc.noaa.gov/eog/download.html). The water treatment capacity of WWTPs were obtained from a national study by Chen et al. (2019).

2.2. Linking Lake N and P Concentrations to Watershed Characteristics

To establish the linkages along the lake-watershed continuum, a hydrological-distance analysis module was developed to extract the boundary and hydrological connectivity for 104 lake-watershed systems (Figure 2). GLMs were developed to predict lake N and P concentrations based on watershed features within various (<1, <2, …, <50 km) hydrological distances, and subsequently estimate nutrient attenuation rates from watersheds to lakes.

2.2.1. Drainage Networks With Flow Direction to Characterize Lake-Watershed Systems

A hydrological-distance analysis module (HDA) was developed to identify the watershed boundary (upstream area) of each lake, and to establish drainage network with flow direction that characterize mass exchange along the lake-watershed continuum.

The module was implemented using the Python package of GDAL and PCRaster. Compared with the existing vector data set for lakes and watersheds, such as the Lake-Catchment data set (Hill et al., 2018) and the data set of HydroLAKES (Messager et al., 2016), the HDA module is targeted at the drainage direction within a watershed rather than watershed boundary. Based on the HDA module, we characterized the watershed boundary and hydrological distances for 104 lake-watershed systems according to the following steps: (a) A raster map with a

500×500 m resolution was derived using lake boundary data. (b) A drainage direction network, covering the lake and its surrounding areas, was developed using an eight-flow direction matrix (D8) from the PCRaster package. (c) Nearshore cells of the lake were classified as sub-watershed outlet points. Each nearshore cell was linked to a sub-watershed, and each cell within these sub-watersheds was assigned a value representing the hydrological distance of the cell from the corresponding lake. The raster images were transferred into vector files (Shapefile format), and therefore each lake was assigned a watershed boundary file, along with a hydrological distance file including 50 polygons that had a hydrological distance of 1, 2, …, 50 km to the lake. Two examples (Lakes Taihu and Chaohu) of this process can be found in Figure S3 of Supporting Information S1.

2.2.2. Lake-Watershed Models in Relation to Hydrological Distance

Based on the data sets (Section 2.1) and watershed boundaries (Section 2.2.1), we generated paired data of lake N and P concentrations and watershed characteristics for the selected 104 lake-watershed systems. The average values of the watershed characteristics for various areas within a certain (1, 2, …, 50 km) hydrological distance were calculated. GLM models were then developed to predict TP, TN and NH_3-N concentrations using predictors reflecting watershed characteristics within various (<1, <2, …, <50 km) hydrological distances. In their generic form, GLM model equations can be described as follows:

$$\mathrm{WQ}_i^k = \mathrm{GLM}\left(x_1^k,\ldots,x_j^k\ldots,x_n^k\right) \tag{1}$$

where x_j^k represents the predictor variable j ($j = 1, 2, \ldots, 9$) from the area with a hydrological distance of $<k$ km. WQ_i^k ($i = 1, 2, 3$) represents the predicted variable i (TN, NH_3-N, or TP) using the x_j^k predictors as model inputs. Therefore, a total of 50 ($k = 1, 2, \ldots, 50$) GLM models were developed to provide a prediction per lake for nutrient considered.

Nine predictor variables, both natural and anthropogenic, were used to characterize lake N and P variability (Figures S1 and S2 in Supporting Information S1). The natural predictor variables included slope (°) and elevation (m) to represent the capacity of two prevalent landscape features to determine mass transport rates and consequently in-stream attenuation, daily average air temperature (^{o}C) and annual precipitation (mm/yr) to account for the influences of weather conditions on biogeochemical processes and flow rates, watershed-to-lake ratio to represent the areal extent of the contributing land relative to the surface area of the receiving waterbody. Both lake and watershed areas were obtained from the corresponding boundary files. The anthropogenic predictors included area percentage of urban (%) and area percentage of farmland (%) as proxies for non-point source pollution associated with the extent of agriculture and urbanization in the surrounding watersheds, water treatment capacity of WWTPs (t/d/km^2) to discern China's efforts in controlling point-source pollution, nighttime light intensity to collectively represent the magnitude of human footprint. The nighttime light intensity had a high spatio-temporal resolution and was thus more suitable for our national analysis. To address potential collinearity problems among these 9 predictor variables, recursive feature elimination (RFE) was used to determine targeted predictors used in our models (Liu et al., 2021). The predictor selection was subsequently conducted for the development of each GLM (Neter et al., 1996).

2.2.3. Characterization of Nitrogen and Phosphorus Attenuation Rates From Watersheds to Lakes

To further quantify N and P attenuation rates from watersheds to lakes, empirical models were developed based on the assumption that the watershed area with a shorter hydrological distance can exert greater control on lake conditions. The governing equation for this exercise is expressed as:

$$\mathrm{WQ}_i = \mathrm{GLM}\left(\bar{x}_1,\ldots,\bar{x}_j\ldots,\bar{x}_n\right) \tag{2}$$

$$\bar{x}_j = \sum_{k=1}^{50}\left[\max\left(1-\mu_k,0\right)\times k_s \times x_j^k\right] \tag{3}$$

$$\mu_k = \mu_0 \times \theta^k \tag{4}$$

$$k_s = S_k/S_1 \tag{5}$$

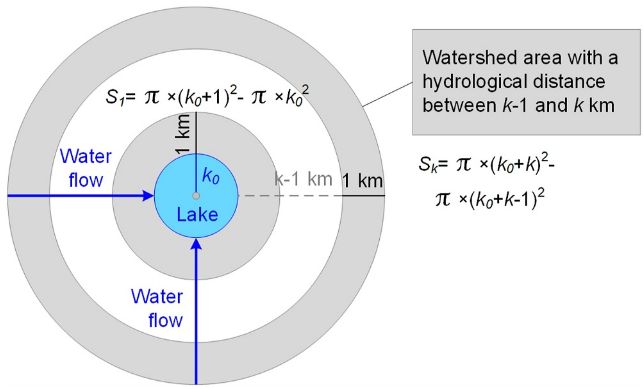

Figure 3. Conceptual description of a lake with the surrounding watershed. The watershed sector with a hydrological distance between k and $k-1$ km has an area of S_k. The sector with a hydrological distance between 0 and 1 km has an area of S_1.

where $WQ_i (i = 1, 2, 3)$ represents the predicted variable i (TN, NH_3-N, or TP) using the relationships developed in the previous step (Equation 1); $\bar{x}_j$ represents a weighted average of the predictor variable j ($j = 1, 2, \ldots, 9$) that is calculated using Equation 3; μ_k is a weight coefficient (0–1) representing the attenuation rate for the area/watershed sector with a hydrological distance between $k-1$ and k km; μ_0 is a decay/attenuation constant, and θ is the attenuation coefficient for the watershed sector k; k_s is the ratio of the watershed area S_k:S_1; S_k represents the watershed area with a hydrological distance between k and $k-1$ km; S_1 represent the watershed area with a hydrological distance between 0 and 1 km. Assuming a circular shape for the watershed surrounding each lake (Figure 3), S_1 and S_k can be empirically estimated by $\pi \times (k_0 + 1)^2 - \pi \times k_0{}^2$ and $\pi \times (k_0 + k)^2 - \pi \times (k_0 + k - 1)^2$, respectively. The contribution of the watershed sector k is determined by weighting the value of each predictor variable j with the corresponding area S_k (larger as we move from the lake to upstream sites) and the hydrological distance from the receiving waterbody (expressed as a power function, in which the exponent is the distance from the lake inlet). We used a random sampling method to generate permutations for the parameters μ_0 and θ that determine the contribution of each hydrological-distance area, after assigning uniform distributions to each parameter within plausible ranges. We generated 2,500 (50×50) parameter sets to obtain the optimal combination that resulted in the best performance for each of the TN, NH_3-N and TP models.

3. Results

3.1. Lake-Watershed Relationships as a Function of the Hydrological Distance

Through the investigation of lake-watershed relationships, we found that N and P concentrations can be accurately estimated by natural and/or anthropogenic characteristics within a certain hydrological distance rather than the entire watershed. Although the model fit against the in-lake TN, NH_3-N and TP levels varied with the hydrological distances from 1 to 50 km, all the 150 ($=50 \times 3$) predictive outputs achieved satisfactory model fit with $P < 0.05$ (Figure 4). TP models displayed superior performance relative to their counterparts for TN and NH_3-N concentrations. Specifically, the predicted lake TP values were closer to the observed ones, $R^2 = 0.52 \pm 0.04$ (mean value ± standard deviation), than those for TN and NH_3-N, for which the R^2 values were 0.17 ± 0.005 and 0.35 ± 0.05, respectively (Figure 4a). The latter R^2 difference may reflect the more complex terrestrial and aquatic biogeochemical processes associated with the nitrogen cycle. Except from the TN models (Figure 4b), the performance (almost) monotonically increased with increasing hydrological distance once a threshold value of 5 km was exceeded. Interestingly, the NH_3-N and TP models had similar response to hydrological distance (Figures 4c and 4d), and achieved the best model fit ($R^2 = 0.41$ and 0.57 respectively) at a hydrological distance of 45 and 44 km, respectively. TN, NH_3-N and TP models leading to the best fit can be found in Section 1 of the Supporting Information S1. The variable selection process consistently identified the water treatment capacity (WWTP) along with the percentage area of farmland (FarmS) as the two most significant predictors for NH_3-N and TP concentrations. The inclusion of the former variable likely weakened the signature of other proxies of the human footprint in urbanized settings, such as the percentage urban area (UrbanS) and the nighttime light intensity (LI). Likewise, the annual precipitation (PR) has a discernible role in shaping in-lake NH_3-N and TP levels, whereas the impact of the watershed elevation (DEM) and/or slope (SL) appears to be relatively minor, and even in some instances, the derived regression coefficients were counterintuitive (see negative relationship between TP concentration and watershed slope).

To further shed light on the capacity of the hydrological distance to modulate the lake-watershed relationships, 104 samples (lakes) were separated into two (high- and low-value) groups according to the median values of the predictors (Figure 1d and Figure S4 in Supporting Information S1). Our comparison revealed that human activities pose challenges in predicting lake N and P concentrations. Furthermore, human activities increased the required hydrological distance to predict lake N and P using the watershed characteristics. Based on the model fit patterns in response to the hydrological distance (Figure 5), we found that the high-value group for the anthropogenic variables generally led to higher performance for both NH_3-N and TP concentrations, when longer

10.1029/2024EF004860

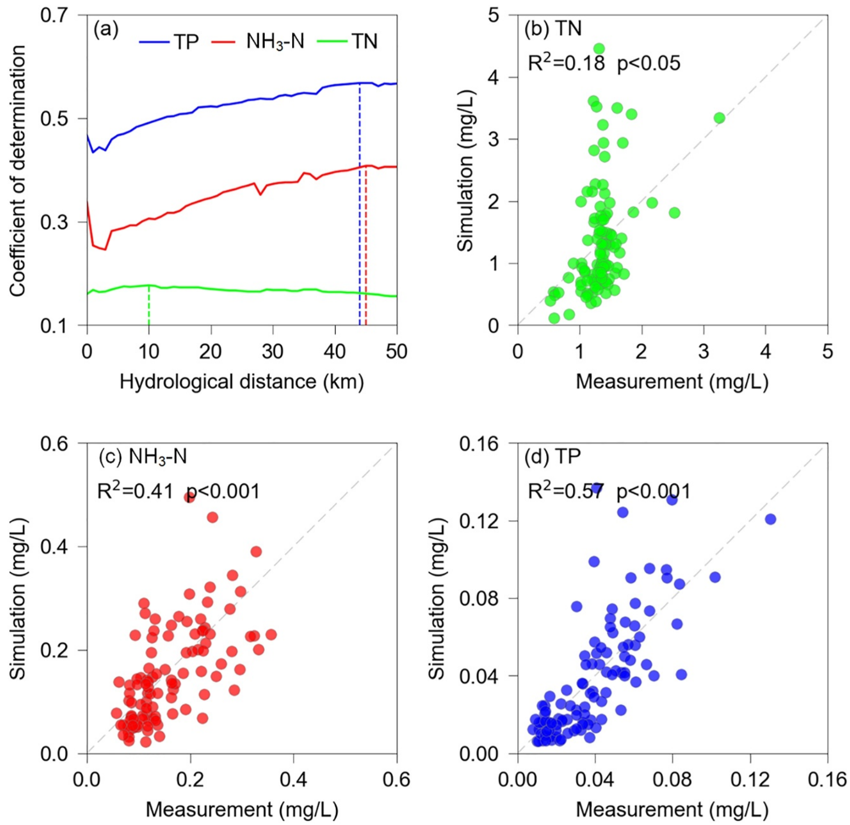

Figure 4. (a) TN, NH_3-N and TP model performance using as predictor variables the watershed characteristics within hydrological distances varying from 1 to 50 km. (b–d) Highest performing models to predict TN, NH_3-N, and TP concentrations within hydrological distances of 10, 45, and 44 km, respectively.

hydrological distances (39.3 ± 7.5 km) were considered. Natural predictor variables did not reveal a consistent impact on lake-watershed relationships, as the model fit between low and high values did not result in consistently superior (or inferior) performance of the lake-watershed relationships. The only notable exception was the catchment slope as a predictor for lake TP, where steeper slopes had improved performance relative to flatter watersheds (Figure 5b). This result probably indirectly reflects the strong relationship between slope and human footprint, in that watershed areas with higher slope generally have less intense anthropogenic activities.

3.2. Nutrient Attenuation From Watersheds to Lakes

Based on 7,500 ($3 \times 50 \times 50$) tested models, we identified the optimal parameter sets ($\mu_0^{TN} = 0.05$ and $\theta = 1.065$; $\mu_0^{NH} = 0.05$ and $\theta = 1.057$; $\mu_0^{TP} = 0.29$ and $\theta = 1.021$) that can be used to infer on the TN, NH_3−N, and TP attenuation rates from watersheds to lakes (Figures 6a–6c). These three models provided the best model fit ($P < 0.001$) with R^2 values of 0.57, 0.10 and 0.41 for TP, TN and NH_3−N, respectively. Based on these parameter estimates, the average attenuation rates within 1 km from the lake inlet were equal to 0.0533, 0.0529 and 0.2961 for TN, NH_3−N, and TP, respectively. In a similar manner, nutrients originating from watershed sectors with hydrological distances of 47, 54, and 60 km can still contribute non-point source loads downstream for the three nutrient forms. The same derived attenuation rates also suggest that the natural and anthropogenic watershed attributes within a hydrological distance of 50 km could explain >40% and nearly 60% of the NH_3−N and TP variability in the receiving waterbodies (Figure 7).

The spatial distribution of the model error with the optimal parameter sets revealed that both N and P for the lakes at the middle and lower reaches of Yangtze River were better predicted, which incidentally is the most intensively studied area (Figure 7). The distribution of prediction bias (= predictive value–measured value) revealed that there is little systematic bias for NH_3−N and TP models, in that there are no geographic regions displaying excessively high error values.

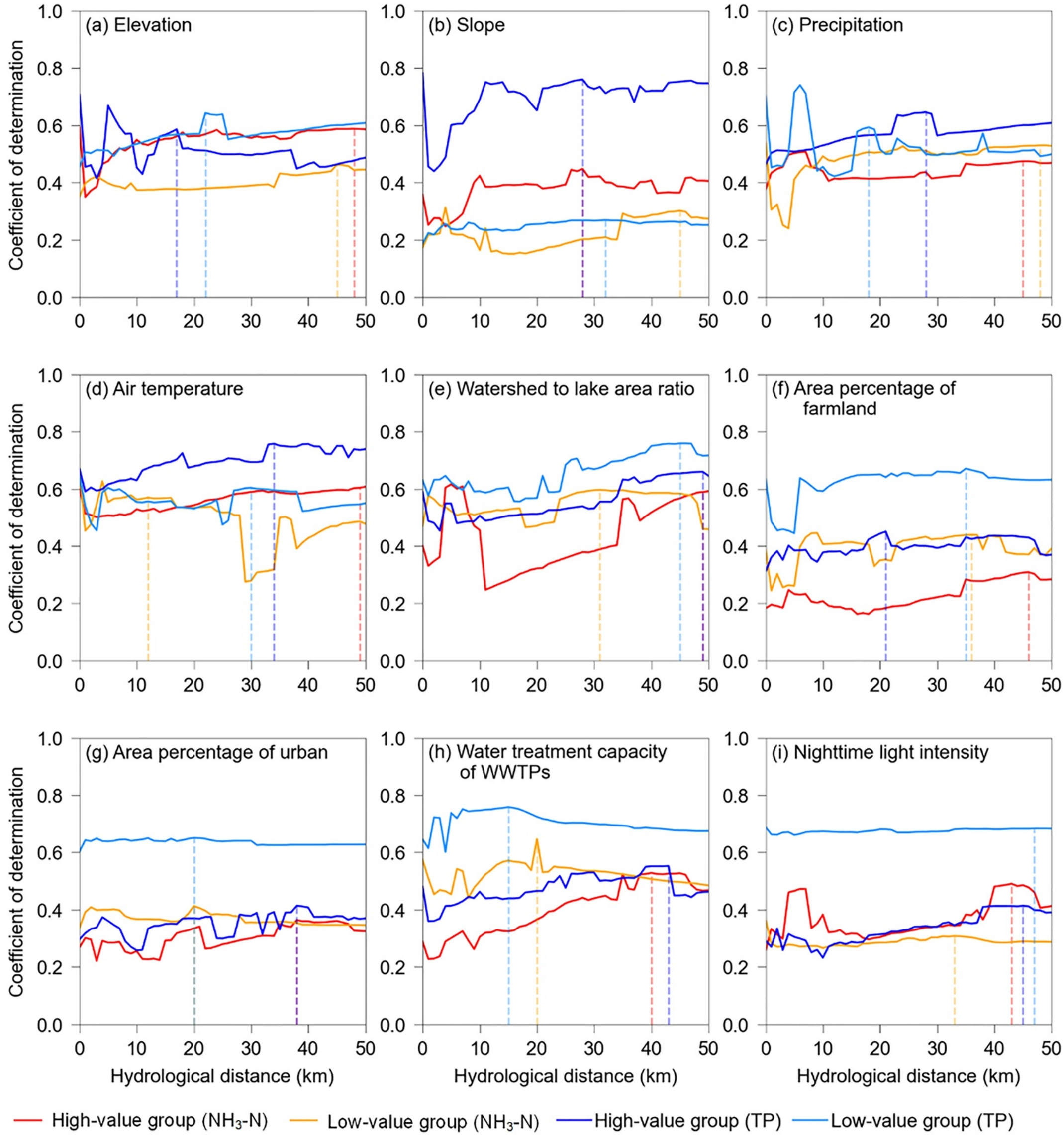

Figure 5. TN, NH_3-N, and TP model fit with varying (1–50 km) hydrological distance for both high- and low-values of each predictor variable. The best model fit within a hydrological distance of 6–50 km for each group is labeled as dash line.

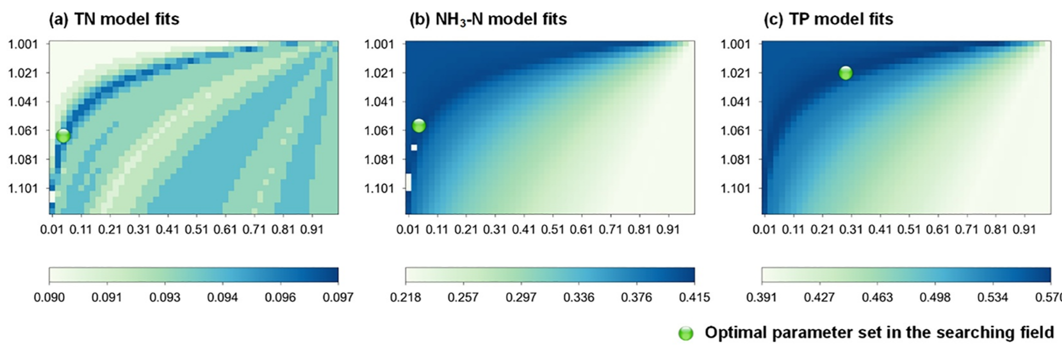

Figure 6. TN, NH_3-N, and TP model fit as a function of the attenuation rates during the transport processes within watershed context, as represented by the parameters of $\boldsymbol{\mu}_0$ (attenuation constant) and $\boldsymbol{\theta}$ (attenuation coefficient).

10.1029/2024EF004860

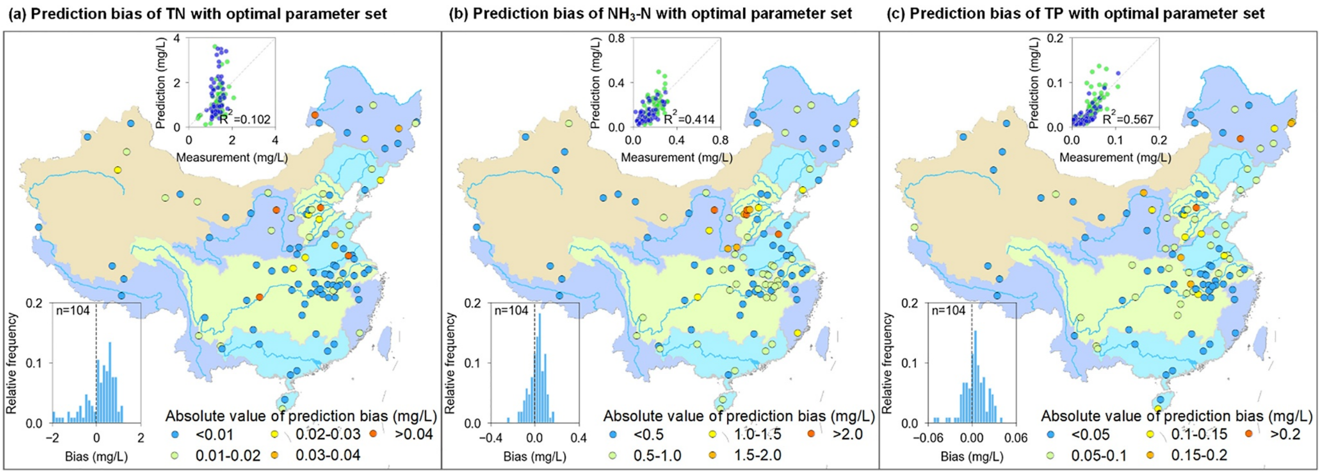

Figure 7. Prediction bias of (a) TN, (b) NH_3-N, and (c) TP concentrations at 104 lakes/reservoirs across China. Insert panels show the error distribution along with the model fit scatterplots based on the lakes (green dots) and reservoirs (blue dots).

4. Discussion

Landscape attributes modulate terrestrial contributions to water paths by shaping the amount of external fluxes along with their fate and transport within the receiving water bodies (Hou et al., 2022; Li et al., 2022; Smits et al., 2021). Consequently, the establishment of relationships between catchment characteristics, land degradation, human-environment interactions, and lake water-quality deterioration has been highly concerned (Wei et al., 2020; Zwart et al., 2018). Most of these empirical models aimed to elucidate the relationships at the macroscale (e.g., regional to national scales), and have been developed using cross-section data sets, which included multiple-site measurements across broader regions. Due to the significance of among-site variability, such data sets often exhibit broader ranges for the variables involved in the empirical relationships, and the subsequent model fit frequently results in well-identified parameter estimates (Cheng et al., 2010; Shimoda & Arhonditsis, 2015). Despite their simplicity in both concept and structure, these macroscale models are seen as important tools in management practices to increase our understanding on the populations of watershed-lake systems, and enhance the ability to extend investigation results to unobserved sites or scale-up local results to larger scales, assist with the recovery of impaired waterbodies, and the conservation of unaffected lakes (Huang et al., 2020b, 2021; Lapierre et al., 2015; Soranno et al., 2015). Within this context, this study presented a national-scale characterization between a set of natural and anthropogenic watershed characteristics and in-lake nutrient levels, augmented by a novel hydrological-distance analysis framework to delineate the watershed sectors with the strongest footprint on lake nutrient dynamics.

Based on a broad review of existing attempts to link lake water quality with watershed characteristics (Table S1 in Supporting Information S1), data-driven models were characterized by a large performance variation, generally evaluated in terms of both fit statistics (e.g., R^2) and evidence of model statistical significance; that is, evidence to falsify the null hypothesis of a lack of a discernible relationship between response and predictor variables. The models presented here are both statistically significant ($P < 0.05$) with moderate to satisfactory performance ($R^2 = 0.18$–0.57). Our models were able to establish acceptable relationships among in-lake TP, land-use patterns, landscape features, and watershed hydrology. The higher model fit of TP compared with TN is plausible. TP concentrations often exhibit a strong relationship with flow, owing to the active transport of particulate constituents either by overland flow or via soil macropores to tile drains, as well as due to their capacity to be readily remobilized from the streambed/bank (Godsey et al., 2009; Meybeck & Moatar, 2012). Contrary to our understanding of P fate and transport, there is less clarity regarding an overall concentration-discharge paradigm for N, primarily because a larger proportion of TN is present in the dissolved phase. Because of their high solubility, nitrite/nitrate can be transported through diverse flow pathways including overland, subsurface or groundwater, and demonstrating apparent stability of the concentrations in relation to flow fluctuations (Long et al., 2014, 2015). Ammonium and total Kjeldahl nitrogen, as compared to other N species, exhibit relatively low solubility and/or lower susceptibility to subsurface leaching due to their immobilization by clay or other soil chemical

10.1029/2024EF004860

constituents, and may thus notably increase during snowmelt and precipitation events (Long et al., 2014). The N cycling within water, soil/sediment, and biota is driven by diverse biogeochemical processes, which contribute to legacy accumulation and long turnover times, such as N uptake by plants and soil microbial community, plant root or litter decomposition, nitrification and denitrification, soil organic matter ammonification, mineral precipitation/dissolution, and sorption/desorption (Beaulieu et al., 2011; Seitzinger et al., 2006). It is thus not surprising that the fit of our TN models was relatively low ($R^2 = 0.18$), which in turn underscores the importance to characterize N cycle-related processes at specific sites (Cui et al., 2013; Liu et al., 2016; Wu et al., 2022).

Another major finding of our national-scale study is that ambient nutrient concentrations depend strongly on watershed characteristics and delivery mechanisms (or attenuation rates) within a specific hydrological distance rather than the entire watershed or the narrower lake buffering zone (Nielsen et al., 2012; Peng et al., 2021; Wollheim et al., 2022). Moreover, depending on the nutrient form examined, our analysis showed that watershed sectors with hydrological distances between 45 and 60 km from the lake shoreline can still contribute nutrients downstream. In the same vein though, recent research has added another layer of complexity by signifying the variation in in-stream attenuation rates that arises from the prevailing flow regimes and stream order/size considered (Neumann, Saber, et al., 2021; Wellen et al., 2012). Specifically, small- and medium-flow streams tend to display higher attenuation rates, and consequently higher nutrient retention. This is due to the higher sedimentation rates and larger biotic uptake in the context of slower flow velocities, shallow stream depths, and consequently longer residence times, particularly in dry years compared to wet years (Aguilera et al., 2012; Behrendt & Opitz, 1999). In contrast, large streams are characterized by lower nutrient retention without significant difference between dry and wet years (Neumann, Saber, et al., 2021). Viewed in this context, the moderately strong signature of precipitation as predictor variable of in-lake nutrient conditions in our models is not surprising. Likewise, the areal extent of farmland and/or the proxies of urbanization (nighttime light intensity and percentage of urban area) had a plausibly discernible signature on the ambient nutrient conditions of the receiving waterbody, especially for the TP and NH_3-N models. In particular, Chinese agriculture has achieved yields that are nearing the maximum attainable for a wide range of crops (Mueller et al., 2012). However, the side-effect of this successful outcome is the excessive use of fertilizers with a rate twice as high to the amount for crop assimilation (Liu et al., 2016). Legacy nutrients stored in soils can then be influenced by various hydrological and biogeochemical mechanisms (Basu et al., 2010), which modulate the magnitude and delivery rate of nutrient loading downstream. Consistent with our results, a significant body of literature also emphasized the importance of urban runoff as a primary source of nutrients, surpassing other land uses, but also as the non-point source exhibiting the greatest discrepancies in their impact between dry and wet years (Neumann, Saber, et al., 2021). The increased nutrient loading originating from urban areas can be attributed to the flashiness of urban watersheds caused by higher hydrological connectivity to drainage networks, and the heightened likelihood of precipitation events generating surface runoff and the resulting loading of constituents in impermeable areas (Bieroza et al., 2018).

While the notion that water quality is intricately determined by the attributes of the surrounding watersheds is well established in the limnological literature (Paltsev & Creed, 2022; Peng et al., 2021; Wu & Lu, 2021), identifying hot spots of nutrient export along the watershed-lake continuum arguably remains a controversial topic in the context freshwater management. Recognizing the need to mitigate the ubiquitous water-pollution problems, the Chinese government released the Lake Chief System, which represents a roadmap on the land-use planning and design of the nearshore zone. However, the extent of the nearshore area, where restoration measures for nutrient control should be prioritized, is not clearly defined and may be a contentious issue. For example, water quality managers in Lake Taihu treated the nearshore area within a 3-km distance from the shoreline, whereas the local government used a 5-km buffer distance. Even more so, our analysis showed that the distance of the watershed sectors with a discernible impact on ambient nutrient concentrations can be even longer (14–45 km in Figure 5). This lack of clarity in the definition of the nearshore area or the variability in the catchment size that can conceivably impact in-lake conditions hinders the design of effective nutrient control strategies in space and time. Our study provided an avenue to determine the critical hydrological distance in any given watershed by establishing linkages between ambient nutrient variables (e.g., TN, NH_3-N and TP) and watershed attributes associated with natural factors and anthropogenic activities alike. This more granular characterization of the effective hydrological-distance, instead of expressing the predictor variables over the entire watershed, can offer a first approximation of the urban and agricultural sites that should be prioritized for the implementation of best management practices (BMPs) in a cost-effective and ecologically beneficial manner (Eckart et al., 2017; Martin-Mikle et al., 2015).

In China, the severe eutrophication problems in Lakes Taihu and Chaohu have received considerable public attention due to the harmful impact to a large population (Huang, Zhang, et al., 2020). To address these problems, the government has proposed a series of management strategies (Table S3 in Supporting Information S1) that heavily rely on the determination of the boundaries of the nearshore areas as well as the priority sites within that zone. The nearshore areas of the two lakes comprise both highly developed urbanized sites and intensive agricultural land (Figures S1c and S1d in Supporting Information S1), including aquaculture ponds and rice-wheat rotation farmlands. Such intensive agricultural activities are particularly prone to high N and P losses, and are thus conducive to exacerbating lake eutrophication. Remedial measures to mitigate the eutrophication severity involve the optimization of fertilizer use, seasonal land abandonment for agriculture farmlands, wastewater recycling for aquaculture ponds, and construction of WWTPs (Huang, Chen, et al., 2020; Jiangsu Provincial Department of Ecology and Environment, 2017; Yan et al., 2022). However, cost-efficiency of these actions has been debated, mainly due to the lack of scientific evidence in establishing precision farming practices, such as "4R" fertilization strategies (right time, right source, right rate, and right placement of fertilizers). The traditional shoreline-buffering approach has an implicit assumption that the areas closer to the watercourses exert a greater water-quality control, for which our analysis showed may not necessarily be true as the areas with discernible impact can be located far from the vicinity of a given lake (Figures S1a and S1b in Supporting Information S1). The introduction of the hydrological distance into our simple empirical models can be regarded as a relative weight in calculating the pollutant contribution from a specific watershed sector (Figure 7), and can thus offer insights into the candidate sites that should be prioritized for the implementation of mitigation measures.

From a modeling standpoint, the most popular practice to estimate the magnitude of non-point source pollution is based on sophisticated spatially distributed process-based models. Their capacity to partition watersheds into multiple distinct units to capture the spatial heterogeneity of parameters and inputs makes them highly relevant for source attribution exercises and assessment of land-use management scenarios (Wellen et al., 2015). Nevertheless, a common approach using distributed process-based models involves their calibrating them against a single point, typical a gauging site located at the outlet of the watershed (Wellen et al., 2015), which could undermine their capacity to accommodate the spatial heterogeneity on topography, land use, soil characteristics, and human activities within a watershed (Dong et al., 2019; Nossent & Bauwens, 2012). The use of over-parameterized process-based models can lead to a convoluted exercise that is confounded by significant uncertainty. Thus, the decision-making processes can be strengthened by coupling them with data-driven (or empirical) watershed models, like those presented in our analysis (Neumann, Dong, et al., 2021). Alongside their capacity to offer first approximations to characterize critical watershed processes, using parsimonious data-driven models can provide a statistically rigorous means to pinpoint unnoticed or newly emerged nutrient "hot spots" or "hot moments", as well as critical pathways for nutrient transport (Kovacs et al., 2012; Long et al., 2015). The outcomes produced by data-driven watershed models can also be utilized to highlight the need to improve subroutines of complex mechanistic models, and to obtain the required data to calibrate, validate and verify them effectively (Arhonditsis et al., 2019a, 2019b). Thus, our empirical lake-watershed modeling framework not only can support a regional assessment for high risk of nutrient export in time and space, but is also conducive to identifying areas to advance our mechanistic understanding at the catchment level (Neumann, Dong, et al., 2021). However, our simple empirical models have limited capacity to represent the multitude of hydrological and biogeochemical processes (surface runoff, groundwater, sediment transport, nutrient cycling, and channel routing) that determine the fate and transport of nutrients within the studied lake-watershed ecosystems. To overcome this weakness, the establishment of model ensembles, coupling data-driven (e.g., statistical and machine learning) together with mechanistic models, offers an appealing strategy that can address a wide range of conceptual and operational uncertainties typically underlying any modeling exercise (Reichstein et al., 2019).

5. Conclusions

A variety of management practices are implemented to mitigate the impact of nonpoint source pollution assuming their effectiveness is guaranteed in short and long term. However, emerging evidence indicates significant variability in their outcomes and moderate improvements in nutrient conditions downstream, often considerably lower than the specific requirements of their expectation. The present study aimed to establish a macroscale modeling framework that will ultimately connects regional scale ("big picture") analysis with watershed-level design of management plans of non-point sources at individual sites in agricultural and urban settings. Our national-scale investigation revealed that China's lake nutrient concentrations can be satisfactorily predicted by watershed attributes, and was particularly sensitive to proxies of natural drivers and anthropogenic activities within a <45 km hydrological distance. The strength of the human footprint has an impact of the predictive capacity, as well as the hydrological distance within a watershed context that can be relevant for predicting lake nutrient levels. The design of optimal nutrient-control practices has become a cornerstone of the on-going water quality remediation measures in China. Therefore, the development of a parsimonious spatially explicit, network of empirical modeling tools is essential to identify high-risk areas with elevated pollutant export, and determine land-to-water transport and in-stream attenuation across a multitude of watersheds in China. Such modeling tools and their application would enhance our understanding of how land uses and extreme events impact flow regimes and biogeochemical cycling within a watershed.

Data Availability Statement

The collected data set for our modelling practice is provided in Table S4 of Supporting Information S1 and available in Huang et al. (2024).

Acknowledgments
This work was supported by National Natural Science Foundation of China (42222104, 42371116), the Science and Technology Innovation Program of Carbon-emission-peak and Carbon-neutrality in Jiangsu Province (BK20220042), Science and Technology Planning Project of NIGLAS (NIGLAS2022GS10). Additional funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant (#06283–2017 to G. Arhonditsis). Hans W. Paerl was supported by the US National Science Foundation (1831096 and 2108917) and the National Institutes of Health (1P01ES028939-01). The authors would like to thank Jing Zhang (Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences), Rui Qian and Yinghao Qu (Hohai University, China) for improving figures. The authors would like to thank the National Tibetan Plateau Data Center for providing the meteorological data for our lake-watershed modelling.

References

Aguilera, R., Marce, R., & Sabater, S. (2012). Linking in-stream nutrient flux to land use and inter-annual hydrological variability at the watershed scale. *Science of the Total Environment*, *440*, 72–81. https://doi.org/10.1016/j.scitotenv.2012.08.030

Arhonditsis, G. B., Neumann, A., Shimoda, Y., Kim, D.-K., Dong, F., Onandia, G., et al. (2019a). Castles built on sand or predictive limnology in action? Part A: Evaluation of an integrated modelling framework to guide adaptive management implementation in Lake Erie. *Ecological Informatics*, *53*, 100968. https://doi.org/10.1016/j.ecoinf.2019.05.014

Arhonditsis, G. B., Neumann, A., Shimoda, Y., Kim, D.-K., Dong, F., Onandia, G., et al. (2019b). Castles built on sand or predictive limnology in action? Part B: Designing the next monitoring-modelling-assessment cycle of adaptive management in Lake Erie. *Ecological Informatics*, *53*, 100969. https://doi.org/10.1016/j.ecoinf.2019.05.015

Arnillas, C. A., Yang, C., Zamaria, S. A., Neumann, A., Javed, A., Shimoda, Y., et al. (2021). Integrating watershed and ecosystem service models to assess best management practice efficiency: Guidelines for Lake Erie managers and watershed modellers. *Environmental Review*, *29*(1), 31–63. https://doi.org/10.1139/er-2020-0071

Basu, N. B., Destouni, G., Jawitz, J. W., Thompson, S. E., Loukinova, N. V., Darracq, A., et al. (2010). Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. *Geophysical Research Letters*, *37*(23), L23404. https://doi.org/10.1029/2010gl045168

Beaulieu, J. J., Tank, J. L., Hamilton, S. K., Wollheim, W. M., Hall, R. O., Mulholland, P. J., et al. (2011). Nitrous oxide emission from denitrification in stream and river networks. *Proceedings of the National Academy of Sciences of the U S A*, *108*(1), 214–219. https://doi.org/10.1073/pnas.1011464108

Behrendt, H., & Opitz, D. (1999). Retention of nutrients in river systems: Dependence on specific runoff and hydraulic load. *Hydrobiologia*, *410*, 111–122. https://doi.org/10.1023/a:1003735225869

Beusen, A. H. W., Bouwman, A. F., Van Beek, L. P. H., Mogollón, J. M., & Middelburg, J. J. (2016). Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. *Biogeosciences*, *13*(8), 2441–2451. https://doi.org/10.5194/bg-13-2441-2016

Beusen, A. H. W., Doelman, J. C., Van Beek, L. P. H., Van Puijenbroek, P. J. T. M., Mogollón, J. M., Van Grinsven, H. J. M., et al. (2022). Exploring river nitrogen and phosphorus loading and export to global coastal waters in the Shared Socio-economic pathways. *Global Environmental Change*, *72*, 102426. https://doi.org/10.1016/j.gloenvcha.2021.102426

Bieroza, M. Z., Heathwaite, A. L., Bechmann, M., Kyllmar, K., & Jordan, P. (2018). The concentration-discharge slope as a tool for water quality management. *Science of the Total Environment*, *630*, 738–749. https://doi.org/10.1016/j.scitotenv.2018.02.256

Chen, X., Strokal, M., Van Vliet, M. T. H., Stuiver, J., Wang, M., Bai, Z., et al. (2019). Multi-scale modeling of nutrient pollution in the rivers of China. *Environmental Science & Technology*, *53*(16), 9614–9625. https://doi.org/10.1021/acs.est.8b07352

Cheng, V., Arhonditsis, G. B., & Brett, M. T. (2010). A revaluation of lake-phosphorus loading models using a Bayesian hierarchical framework. *Ecological Research*, *25*(1), 59–76. https://doi.org/10.1007/s11284-009-0630-5

Cui, S., Shi, Y., Groffman, P. M., Schlesinger, W. H., & Zhu, Y. (2013). Centennial-scale analysis of the creation and fate of reactive nitrogen in China (1910–2010). *Proceedings of the National Academy of Sciences*, *110*(6), 2052–2057. https://doi.org/10.1073/pnas.1221638110

Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., et al. (2018). Pursuing sustainable productivity with millions of smallholder farmers. *Nature*, *555*(7696), 363–366. https://doi.org/10.1038/nature25785

Dong, F., Neumann, A., Kim, D.-K., Huang, J., & Arhonditsis, G. B. (2019). A season-specific, multi-site calibration strategy to study the hydrological cycle and the impact of extreme-flow events along an urban-to-agricultural gradient. *Ecological Informatics*, *54*, 100993. https://doi.org/10.1016/j.ecoinf.2019.100993

Eckart, K., McPhee, Z., & Bolisetti, T. (2017). Performance and implementation of low impact development – a review. *Science of the Total Environment*, *607–608*, 413–432. https://doi.org/10.1016/j.scitotenv.2017.06.254

Godsey, S. E., Kirchner, J. W., & Clow, D. W. (2009). Concentration-discharge relationships reflect chemostatic characteristics of US catchments. *Hydrological Processes*, *23*(13), 1844–1864. https://doi.org/10.1002/hyp.7315

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., & Li, X. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data*, *7*(1), 25. https://doi.org/10.1038/s41597-020-0369-y

Hill, R. A., Weber, M. H., Debbout, R. M., Leibowitz, S. G., & Olsen, A. R. (2018). The Lake-Catchment (LakeCat) dataset: Characterizing landscape features for lake basins within the conterminous USA. *Freshwater Science*, *37*(2), 208–221. https://doi.org/10.1086/697966

Ho, J. C., Michalak, A. M., & Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, *574*(7780), 667–670. https://doi.org/10.1038/s41586-019-1648-7

Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., et al. (2022). Global mapping reveals increase in lacustrine algal blooms over the past decade. *Nature Geoscience*, *15*(2), 130–134. https://doi.org/10.1038/s41561-021-00887-x

Huang, J., Arhonditsis, G. B., Zhang, Y., Zhang, S., Ji, Y., Paerl, H. W., et al. (2024). Dataset to develop the Generalized Linear Models in "The critical role of hydrological distance in shaping nutrient dynamics along the watershed-lake continuum [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.13788260

Huang, J., Chen, Q., Peng, J., & Gao, J. (2020). Quantifying the cost-effectiveness of nutrient-removal strategies for a lowland rural watershed: Insights from process-based modeling. *Ecological Modelling*, *431*, 109123. https://doi.org/10.1016/j.ecolmodel.2020.109123

Huang, J., Zhang, Y., Arhonditsis, G. B., Gao, J., Chen, Q., & Peng, J. (2020). The magnitude and drivers of harmful algal blooms in China's lakes and reservoirs: A national-scale characterization. *Water Research*, *181*, 115902. https://doi.org/10.1016/j.watres.2020.115902

Huang, J., Zhang, Y., Arhonditsis, G. B., Gao, J., Chen, Q., Wu, N., et al. (2019). How successful are the restoration efforts of China's lakes and reservoirs? *Environment International*, *123*, 96–103. https://doi.org/10.1016/j.envint.2018.11.048

Huang, J., Zhang, Y., Bing, H., Peng, J., Dong, F., Gao, J., & Arhonditsis, G. B. (2021). Characterizing the river water quality in China: Recent progress and on-going challenges. *Water Research*, *201*, 117309. https://doi.org/10.1016/j.watres.2021.117309

Huang, J., Zhang, Y., Huang, Q., & Gao, J. (2018). When and where to reduce nutrient for controlling harmful algal blooms in large Eutrophic Lake Chaohu, China? *Ecological Indicators*, *89*, 808–817. https://doi.org/10.1016/j.ecolind.2018.01.056

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471–483. https://doi.org/10.1038/s41579-018-0040-1

Hundey, E. J., Russell, S. D., Longstaffe, F. J., & Moser, K. A. (2016). Agriculture causes nitrate fertilization of remote alpine lakes. *Nature Communications*, *7*(1), 10571. https://doi.org/10.1038/ncomms10571

Ibáñez, C., & Peñuelas, J. (2019). Changing nutrients, changing rivers. *Science*, *365*(6454), 637–638. https://doi.org/10.1126/science.aay2723

Jane, S. F., Hansen, G. J. A., Kraemer, B. M., Leavitt, P. R., Mincer, J. L., North, R. L., et al. (2021). Widespread deoxygenation of temperate lakes. *Nature*, *594*(7861), 66–70. https://doi.org/10.1038/s41586-021-03550-y

Janssen, A. B. G., Janse, J. H., Beusen, A. H. W., Chang, M., Harrison, J. A., Huttunen, I., et al. (2019). How to model algal blooms in any lake on Earth. *Current Opinion in Environmental Sustainability*, *36*, 1–10. https://doi.org/10.1016/j.cosust.2018.09.001

Jiangsu Provincial Department of Ecology and Environment. (2017). Jiangsu provincial department of Ecology and environment. Retrieved from http://www.jiangsu.gov.cn/art/2022/7/7/art_59167_10532559.html

Kovacs, A., Honti, M., Zessner, M., Eder, A., Clement, A., & Bloschl, G. (2012). Identification of phosphorus emission hotspots in agricultural catchments. *Science of the Total Environment*, *433*, 74–88. https://doi.org/10.1016/j.scitotenv.2012.06.024

Lapierre, J. F., Seekell, D. A., & del Giorgio, P. A. (2015). Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. *Global Change Biology*, *21*(12), 4425–4435. https://doi.org/10.1111/gcb.13031

Li, X., Cheng, G., Fu, B., Xia, J., Zhang, L., Yang, D., et al. (2022). Linking critical zone with watershed science: The example of the Heihe River basin. *Earth's Future*, *10*(11), e2022EF002966. https://doi.org/10.1029/2022EF002966

Liu, H., Li, Q., Bai, Y., Yang, C., Wang, J., Zhou, Q., et al. (2021). Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. *Remote Sensing of Environment*, *256*, 112316. https://doi.org/10.1016/j.rse.2021.112316

Liu, X., Sheng, H., Jiang, S., Yuan, Z., Zhang, C., & Elser, J. J. (2016). Intensification of phosphorus cycling in China since the 1600s. *Proceedings of the National Academy of Sciences*, *113*(10), 2609–2614. https://doi.org/10.1073/pnas.1519554113

Long, T. Y., Wellen, C., Arhonditsis, G., & Boyd, D. (2014). Evaluation of stormwater and snowmelt inputs, land use and seasonality on nutrient dynamics in the watersheds of Hamilton Harbour, Ontario, Canada. *Journal of Great Lakes Research*, *40*(4), 964–979. https://doi.org/10.1016/j.jglr.2014.09.017

Long, T. Y., Wellen, C., Arhonditsis, G., Boyd, D., Mohamed, M., & O'Connor, K. (2015). Estimation of tributary total phosphorus loads to Hamilton Harbour, Ontario, Canada, using a series of regression equations. *Journal of Great Lakes Research*, *41*(3), 780–793. https://doi.org/10.1016/j.jglr.2015.04.001

Lu, C. Q., & Tian, H. Q. (2017). Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: Shifted hot spots and nutrient imbalance. *Earth System Science Data*, *9*(1), 181–192. https://doi.org/10.5194/essd-9-181-2017

Ma, R., Yang, G., Duan, H., Jiang, J., Wang, S., Feng, X., et al. (2011). China's lakes at present: Number, area and spatial distribution. *Science China Earth Sciences*, *54*(2), 283–289. https://doi.org/10.1007/s11430-010-4052-6

Martin-Mikle, C. J., de Beurs, K. M., Julian, J. P., & Mayer, P. M. (2015). Identifying priority sites for Low Impact Development (LID) in a mixed-use watershed. *Landscape and Urban Planning*, *140*, 29–41. https://doi.org/10.1016/j.landurbplan.2015.04.002

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 13603. https://doi.org/10.1038/ncomms13603

Meybeck, M., & Moatar, F. (2012). Daily variability of river concentrations and fluxes: Indicators based on the segmentation of the rating curve. *Hydrological Processes*, *26*(8), 1188–1207. https://doi.org/10.1002/hyp.8211

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature*, *490*(7419), 254–257. https://doi.org/10.1038/nature11420

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.), The McGraw-Hill Companies, Inc.

Neumann, A., Dong, F. F., Shimoda, Y., Arnillas, C. A., Javed, A., Yang, C., et al. (2021). A review of the current state of process-based and data-driven modelling: Guidelines for Lake Erie managers and watershed modellers. *Environmental Review*, *29*(4), 443–490. https://doi.org/10.1139/er-2020-0070

Neumann, A., Saber, A., Arnillas, C. A., Shimoda, Y., Yang, C., Javed, A., et al. (2021). Implementation of a watershed modelling framework to support adaptive management in the Canadian side of the Lake Erie basin. *Ecological Informatics*, *66*, 101444. https://doi.org/10.1016/j.ecoinf.2021.101444

Nielsen, A., Trolle, D., Sondergaard, M., Lauridsen, T. L., Bjerring, R., Olesen, J. E., & Jeppesen, E. (2012). Watershed land use effects on lake water quality in Denmark. *Ecological Applications*, *22*(4), 1187–1200. https://doi.org/10.1890/11-1831.1

Nossent, J., & Bauwens, W. (2012). Multi-variable sensitivity and identifiability analysis for a complex environmental model in view of integrated water quantity and water quality modeling. *Water Science and Technology*, *65*(3), 539–549. https://doi.org/10.2166/wst.2012.884

Paltsev, A., & Creed, I. F. (2022). Multi-decadal changes in phytoplankton biomass in northern temperate lakes as seen through the prism of landscape properties. *Global Change Biology*, *28*(7), 2272–2285. https://doi.org/10.1111/gcb.16079

Peng, X., Zhang, L., Li, Y., Lin, Q., He, C., Huang, S., et al. (2021). The changing characteristics of phytoplankton community and biomass in subtropical shallow lakes: Coupling effects of land use patterns and lake morphology. *Water Research*, *200*, 117235. https://doi.org/10.1016/j.watres.2021.117235

10.1029/2024EF004860

Qin, B., Paerl, H. W., Brookes, J. D., Liu, J., Jeppesen, E., Zhu, G., et al. (2019). Why Lake Taihu continues to be plagued with cyanobacterial blooms through 10 years (2007–2017) efforts. *Science Bulletin*, *64*(6), 354–356. https://doi.org/10.1016/j.scib.2019.02.008

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Seitzinger, S., Harrison, J. A., Bohlke, J. K., Bouwman, A. F., Lowrance, R., Peterson, B., et al. (2006). Denitrification across landscapes and waterscapes: A synthesis. *Ecological Applications*, *16*(6), 2064–2090. https://doi.org/10.1890/1051-0761(2006)016[2064:DALAWA]2.0.CO;2

Sharples, J., Middelburg, J. J., Fennel, K., & Jickells, T. D. (2017). What proportion of riverine nutrients reaches the open ocean? *Global Biogeochemical Cycles*, *31*(1), 39–58. https://doi.org/10.1002/2016GB005483

Shimoda, Y., & Arhonditsis, G. B. (2015). Integrating hierarchical Bayes with phosphorus loading modelling. *Ecological Informatics*, *29*, 77–91. https://doi.org/10.1016/j.ecoinf.2015.07.005

Sinha, E., Michalak, A. M., & Balaji, V. (2017). Eutrophication will increase during the 21st century as a result of precipitation changes. *Science*, *357*(6349), 405–408. https://doi.org/10.1126/science.aan2409

Smits, A. P., Currinder, B., Framsted, N., Loken, L. C., Lucero, D., Neal, K. A., et al. (2021). Watershed and lake attributes dictate landscape patterns of resource flows in mountain lakes. *Water Resources Research*, *57*(4), e2020WR027673. https://doi.org/10.1029/2020WR027673

Soranno, P. A., Cheruvelil, K. S., Wagner, T., Webster, K. E., & Bremigan, M. T. (2015). Effects of land use on lake nutrients: The importance of scale, hydrologic connectivity, and region. *PLoS One*, *10*(8), e0135454. https://doi.org/10.1371/journal.pone.0135454

Tong, Y., Wang, M., Peñuelas, J., Liu, X., Paerl, H. W., Elser, J. J., et al. (2020). Improvement in municipal wastewater treatment alters lake nitrogen to phosphorus ratios in populated regions. *Proceedings of the National Academy of Sciences*, *117*(21), 11566–11572. https://doi.org/10.1073/pnas.1920759117

Wei, W., Gao, Y., Huang, J., & Gao, J. (2020). Exploring the effect of basin land degradation on lake and reservoir water quality in China. *Journal of Cleaner Production*, *268*, 122249. https://doi.org/10.1016/j.jclepro.2020.122249

Wellen, C., Arhonditsis, G. B., Labencki, T., & Boyd, D. (2012). A Bayesian methodological framework for accommodating interannual variability of nutrient loading with the SPARROW model. *Water Resources Research*, *48*(10), W10505. https://doi.org/10.1029/2012wr011821

Wellen, C., Kamran-Disfani, A.-R., & Arhonditsis, G. B. (2015). Evaluation of the current state of distributed watershed nutrient water quality modeling. *Environmental Science & Technology*, *49*(6), 3278–3290. https://doi.org/10.1021/es5049557

Wollheim, W. M., Harms, T. K., Robison, A. L., Koenig, L. E., Helton, A. M., Song, C., et al. (2022). Superlinear scaling of riverine biogeochemical function with watershed size. *Nature Communications*, *13*(1), 1230. https://doi.org/10.1038/s41467-022-28630-z

Wu, J., & Lu, J. (2021). Spatial scale effects of landscape metrics on stream water quality and their seasonal changes. *Water Research*, *191*, 116811. https://doi.org/10.1016/j.watres.2021.116811

Wu, Z., Li, J., Sun, Y., Peñuelas, J., Huang, J., Sardans, J., et al. (2022). Imbalance of global nutrient cycles exacerbated by the greater retention of phosphorus over nitrogen in lakes. *Nature Geoscience*, *15*(6), 464–468. https://doi.org/10.1038/s41561-022-00958-7

Wu, Z., Liu, Y., Liang, Z., Wu, S., & Guo, H. (2017). Internal cycling, not external loading, decides the nutrient limitation in eutrophic lake: A dynamic model with temporal Bayesian Hierarchical inference. *Water Research*, *116*, 231–240. https://doi.org/10.1016/j.watres.2017.03.039

Yan, R. H., Li, L. L., Gao, J. F., & Huang, J. C. (2022). Exploring the influence of seasonal cropland abandonment on evapotranspiration and water resources in the humid lowland region, Southern China. *Water Resources Research*, *58*(4), e2021WR031888. https://doi.org/10.1029/2021wr031888

Yin, H., Zhang, M., Yin, P., & Li, J. (2022). Characterization of internal phosphorus loading in the sediment of a large Eutrophic Lake (Lake Taihu, China). *Water Research*, *225*, 119125. https://doi.org/10.1016/j.watres.2022.119125

Yuan, H., Wang, H., Zhou, Y., Jia, B., Yu, J., Cai, Y., et al. (2021). Water-level fluctuations regulate the availability and diffusion kinetics process of phosphorus at lake water–sediment interface. *Water Research*, *200*, 117258. https://doi.org/10.1016/j.watres.2021.117258

Zhang, F., Chen, X., & Vitousek, P. (2013). An experiment for the world. *Nature*, *497*(7447), 33–35. https://doi.org/10.1038/497033a

Zhang, G., Yao, T., Chen, W., Zheng, G., Shum, C. K., Yang, K., et al. (2019). Regional differences of lake evolution across China during 1960s–2015 and its natural and anthropogenic causes. *Remote Sensing of Environment*, *221*, 386–404. https://doi.org/10.1016/j.rse.2018.11.038

Zwart, J. A., Hanson, Z. J., Vanderwall, J., Bolster, D., Hamlet, A., & Jones, S. E. (2018). Spatially explicit, regional-scale simulation of lake carbon fluxes. *Global Biogeochemical Cycles*, *32*(9), 1276–1293. https://doi.org/10.1002/2017gb005843