Wu, Zhen; Li, Jincheng; Sun, Yanxin; [et al.]. «Imbalance of global nutrient cycles exacerbated by the greater retention of phosphorus over nitrogen in lakes». Nature Geoscience, Vol. 15 (June 2022), p. 464-468. DOI 10.1038/s41561-022-00958-7

Imbalance of global nutrient cycles exacerbated by retention of phosphorus in lakes

Zhen Wu[1,2], Jincheng Li[1], Yanxin Sun[1], Josep Penuelas[3,4], Jilin Huang[1], Jordi Sardans[3,4], Qingsong Jiang[1], Jacques C. Finlay[5], Gregory L. Britten[2], Michael J. Follows[2], Wei Gao[6], Boqiang Qin[7], Jinren Ni[1], Shouliang Huo[8], and Yong Liu[1,*]

Affiliations:

[1]College of Environmental Sciences and Engineering, State Environmental Protection Key Laboratory of All Materials Flux in Rivers, Peking University, Beijing, China

[2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

[3]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Catalonia, Spain

[4]CREAF, Cerdanyola del Valles, Catalonia, Spain

[5]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota, USA

[6]Institute of Environmental and Ecological Engineering, Guangdong University of Technology, China

[7]State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China

[8]State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Science, Beijing, China

*Corresponding author. Email: yongliu@pku.edu.cn

Abstract

Imbalanced anthropogenic inputs of nitrogen (N) and phosphorus (P) have significantly increased the ratio between N and P globally, degrading ecosystem productivity and environmental quality. Lakes represent a large global nutrient sink, modifying the flow of N and P in the environment. It remains unknown, however, the relative retention of these two nutrients in global lakes and their role in the imbalance of the nutrient cycles. Here we compare the ratio between P and N in inflows and outflows of more than 5000 lakes globally, using a combination of nutrient budget model and generalized linear model, showing that over 80% of global lakes positively retain both N and P, and almost 90% of the lakes show preferential retention of P. The greater retention of P over N leads to a strong elevation in the ratios between N and P in the lake outflow, exacerbating the imbalance of N and P cycles unexpectedly and potentially leading to biodiversity losses within lakes and algal blooms in downstream N-limited coastal zones. The management of N or P in controlling lake eutrophication has long been debated. Our results suggest that eutrophication management that prioritizes the reduction of P in lakes---which causes a further decrease in P in outflows---may unintentionally aggravate N:P imbalances in global ecosystems. Our results also highlight the importance of nutrient retention stoichiometry in global lake management to benefit watershed and regional biogeochemical cycles.

Global anthropogenic nitrogen (N) and phosphorus (P) inputs to the biosphere have increased asymmetrically since the late 1980s due to human population expansion and industrialization[1]. The N:P ratio of global anthropogenic input increased from 19 to 32 (molar basis hereafter), which is higher than the average N:P ratios of different biospheric compartments (15-16 for open ocean, 16-22 for soils, and 4-30 for organisms)[1-3]. More specifically, the average N:P ratios imported into lakes in the United States remains higher than the Redfield Ratio (16:1) since the 1980s[4]. This imbalance between N and P inputs has altered biogeochemical cycles, negatively affecting biodiversity, water and air quality, ecosystem productivity, and human health[5-8]. Mitigating these negative effects requires understanding nutrient cycles in ecosystems, since they greatly modify nutrient concentration and composition. Lakes are strong nutrient sinks, contributing about 20% of global freshwater nutrient retention and reducing the export of anthropogenic nutrient pollution to downstream ecosystems[9]. Meta-analyses have demonstrated temporal changes in global lake N:P stoichiometry, while it remains unknown whether lake retention is a buffer or accelerator of the imbalance of N:P supply ratios globally. There are notorious difficulties in interpreting variations between studies and observations, as well as the complexity of underlying mechanisms which include multiple processes (e.g., internal sediment release, sedimentation, and denitrification) and various factors (e.g. water residence time, oxygen content, trophic state, and the microbial community)[10-12]. It is also unclear the extent to which internal cycling in lakes drives preferential retention of N or P globally. The altered nutrient flows through lakes may exacerbate adverse effects on primary producer diversity and food web functioning in lakes as well as downstream ecosystems[7,13]. In addition, the lake eutrophication management is primarily driven by observed N:P stoichiometry and prioritization of N vs P mitigation is still under debate[14,15] among scientific and decision-making communities. A greater understanding of the controls over nutrient retention is needed because lakes are increasingly impacted by global change, including sustained eutrophication, warming and altered hydrology.

We use a multi-faceted approach to identify specific patterns of lake nutrient retention and investigate the underlying factors that lead to observed preferential behavior across lakes of different size, location, and trophic gradient. Four trophic states are defined by chlorophyll-a concentration, oligotrophic (<2 µg/L), mesotrophic (2-7 µg/L), eutrophic (7-30 µg/L), and hyper-eutrophic (>30 µg/L) [16]. Three datasets are employed in a progressive approach, with 596 lakes from the U.S. NLA2012 database[16] for the development of a mechanistic nutrient budget model within a Bayesian hierarchical framework, 5,622 lakes from Chlorophyll and Water Chemistry database[17] for nutrient cycling imbalance verification, and 1.4 million lakes from HydroLAKES database[18] for global up-scaling (See Methods for details).

General Patterns of Nutrient Retention

We find that stoichiometric shifts in N:P is a common phenomenon, with higher outflow N:P than inflow in 91.3% of the NLA2012 lakes (Fig. 1a). The average outflow N:P (33.19±18.08) is much higher than the Redfield Ratio of 16 and almost twice as high as the inflow (16.61±12.42). A clear upward trend of nutrient inflow and outflow fluxes emerges across the trophic gradient (Fig. 1e&g). Moreover, a steeper increase of N outflow than inflow along the trophic gradient suggests that internal cycling processes of N may have a greater impact at higher trophic states (Fig. 1e). However, the medians of P outflow are consistently lower than the inflow, indicating a stronger retention capacity of P over N (Fig. 1g). Despite the apparent trends of nutrient fluxes, no significant N:P trend is found in inflow nor outflow. However, the medians of outflow N:P are consistently higher than inflow N:P across the trophic gradient (Fig. 1c), which is consistent with the widely observed elevated N:P of anthropogenic nutrient inputs and a general higher retention of P than N in lakes. Furthermore, the differences between the medians of N:P of outflow and inflow increase with trophic states, corroborating a preferential influence of external nutrient enrichment on process affecting nutrient removal and imbalance in lakes.

Contribution of Internal Nutrient Cycling

Our results further demonstrate that internal nutrient cycling in lakes exacerbate the imbalance of N and P cycles in watersheds. The N:P of in-lake enrichment (defined here as the combined effects of sediment release of P and reactive N and N fixation) is higher than that of in-lake depletion (i.e. the combined effects of sedimentation and denitrification) in 76.5% of the studied lakes (Fig. 1b). The imbalanced N:P ratios of these two processes mirror the imbalance of inflow and outflow N:P ratios, together implying preferential retention of P. In addition, a clear upward trend of in-lake enrichment N:P is found in Fig. 1d with increasing nutrient availability, implying more release of N than P with increasing trophic state. This high release of N is not as well recognized as P release in previous studies that emphasize the importance of internal loading of P in eutrophic lakes[19-22]. More interestingly, the variation of in-lake enrichment N:P decreases along the trophic gradient; while the variation of in-lake depletion N:P remains stable, indicating that the patterns of in-lake nutrient enrichment is more sensitive to the changes of trophic state than in-lake depletion. However, the patterns of preferential P retention persist, despite the increase of in-lake enrichment N:P along trophic gradient as in-lake depletion N:P (9.17±5.31) is always lower than in-lake enrichment (15.85±23.02). The fluxes of in-lake nutrient enrichment and depletion increase along the trophic gradient, but do so unequally between N and P (Fig. 1f&h). Similar to Finlay et al.[23], we find that in-lake N depletion is enhanced by external nutrient enrichment but the capacity to retain N starts to decrease with increasingly high levels of eutrophication. The decline of retention capacity is even faster for N compared to P, implying that preferential P retention drives imbalances leading to increased N:P in lakes and lake outflows.

Global Implication of Nutrient Retention

The preferential pattern of nutrient retention and nutrient cycling imbalance is verified using a large dataset of 5,622 lakes (see Methods for details). 81% of the studied lakes positively retain N and P (Quadrant C in Fig. 2) and 87.8% show a pattern of preferential P retention. In order to quantify the effect of external loading on internal nutrient processes, we define *INCI* (internal nutrient cycling intensity) as the ratio of the sum of absolute values of in-lake nutrient enrichment and depletion to external nutrient loading (see Table 1 for equation of *INCI*). High *INCI* suggests a greater contribution of internal nutrient processes to nutrient retention compared to external loading. The numbers of lakes with *INCI* > 1 for both N and P increase along the trophic gradient, while the numbers for N are always lower than those for P (Table 1). The proportion of lakes with positive nutrient retention decrease along the trophic gradient, implying enhanced in-lake nutrient enrichment in lakes of high trophic state (Table 1). The sharp decrease in the proportion of positive N retention lakes supports previous studies where the contribution of in-lake N enrichment is usually higher in eutrophic lakes than in lakes of lower trophic states[22,24]. Although not all the lakes positively retain N or P, the large majority that do may reduce nutrient levels at the landscape scale. Positive nutrient retention in a single lake provides benefits from reducing the nutrient load in the ecosystem, but may also contribute to the imbalance in global and regional N and P cycles. The model results verify that the overall influence of lakes on global N and P cycles is therefore double-edged. Positive outcomes are due to nutrient retention and act to reduce downstream nutrient pollution. Negative consequences caused by increased N:P ratios may lead to reduced lake food-web biodiversity, decreased drinking water quality, and algal blooms in downstream N-limited coastal zones.

We propagate the distributions of internal nutrient fluxes to global scale with a joint distribution of lake areas and trophic states using a Monte Carlo approach. The joint distribution is generated from the Chlorophyll and Water Chemistry database [17] and HydroLAKES[18] of over 1.4 million lakes (see Methods for details). Our global estimates of nutrient retention by lakes are 28.58 Tg/year (25.23, 31.92, 95% confidence interval, hereafter) for N and 8.16 Tg/year (6.71, 9.62) for P (detailed results shown in supplement). The results are comparable with previous studies by Harrison et al.[25] and Seitzinger et al.[26] with a range of 19.7 to 31 Tg/year for in-lake N depletion. The average N:P of global lake nutrient retention is thus 7.76 (molar basis), much lower than the Redfield Ratio of 16, corroborating the global pattern of preferential P retention. Only 53.2% of global lakes positively retain N while 87.4% positively retain P (Fig. 3). The hotspots of N exports are usually eutrophic lakes with high in-lake N enrichment fluxes but relatively low in-lake depletion fluxes, mainly in East and Central Europe, India, Southeast China, the east coast of the United States, Southeast and Southwest Canada, all with a high population density and intense anthropogenic activities (detailed results shown in supplement).

Nutrient Retention-based Perspective for Eutrophication Management

Our results demonstrate that lake nutrient retention weakens with external nutrient enrichment (Fig. 1). Lakes increasingly but asymmetrically release N and P along the trophic gradient, amplifying the impacts of nutrient pollution and imbalance on downstream ecosystems like coastal regions[23] that are more vulnerable to external N enrichment. Global lake eutrophication management strategies, primarily determined by trophic state and observed N:P ratios, have long been under the debate[14,15,27]. We propose that lake management could be guided by insights gained in our study. Relative patterns of N vs. P fluxes, described by the coordinate system in Fig. 2, could be explicitly incorporated with trophic gradient and physical characters (e.g. water residence time and lake depth). Lakes characterized by export of N and P (due to high in-lake enrichment of N and P, Quadrant A, Fig. 2) are at risk of elevated trophic state, and thus require attention or control of both N and P. Lakes with imbalanced N and P retention (Quadrant B and D, Fig. 2) require special attention to nutrient export from the lake. Oligotrophic and mesotrophic lakes in Quadrant C (Fig. 2) that positively retain both N and P represent low risk of eutrophication and may not require nutrient management; while eutrophic and hyper-eutrophic lakes in Quadrant C are likely to require nutrient management based on the analysis of temporal dynamics of internal nutrient cycling derived from nutrient budget models. Overall, past eutrophication restoration prioritizing P reduction benefits many lakes but may further aggravate the N:P imbalance unintentionally; thus, in consideration of global nutrient cycling, we argue that nutrient retention and stoichiometry balance should be prioritized in lake eutrophication management, particularly in hotspot regions as identified in Fig. 3. Our work further emphasizes the importance of retention stoichiometry as an explicit factor in global lake management.

Acknowledgments: We thank Prof. Huaicheng Guo, Dr. Yue Qin and Dr. Ling Cao for helpful discussion. We also thank High-performance Computing Platform of Peking University for providing computing resources. The present work was financially supported by National Natural

Science Foundation of China (42142047 to YL, 51721006 to YL and JN), Simons Foundation Postdoctoral Fellowship (645921 to GLB), Simons Collaboration on Ocean Processes and Ecology (SCOPE 329108 to MJF), Simons Collaboration for Computational Biogeochemical Modeling of marine Ecosystems (CBIOMES 549931 to MJF), Spanish Government Grant (PID2019-110521GB-I00 and PID2020-115770RB-I00 to JP and JS), and Fundación Ramón Areces Grant (ELEMENTAL-CLIMATE to JP and JS).

Author contributions: ZW and YL conceived the study. ZW and GLB developed the Bayesian mechanistic nutrient budget model. JL, QJ and WG calculated NANI and NAPI for the nutrient budget model. YS carried out the GLM analysis and JH performed the global up-scaling analysis. ZW and YL wrote the paper with direct contributions from JP, JS, JCF, GLB, MJF, BQ, JN, SH.

Competing interests: Authors declare that they have no competing interests.

Table 1. Classification of N and P cycling stoichiometry [a]. The ratio of in-lake nutrient enrichment to depletion less than 1 indicates positive net nutrient retention. The numbers represent the proportions of lakes in each trophic state.

Trophic States	**N**		**P**	
	$EN_N/DE_N < 1$	$INCI > 1$ [b]	$EN_P/DE_P < 1$	$INCI > 1$
Oligotrophic	94.88%	45.59%	99.67%	46.90%
Mesotrophic	83.50%	48.38%	90.88%	56.92%
Eutrophic	73.83%	60.38%	85.63%	66.87%
Hyper-eutrophic	56.46%	69.82%	79.82%	75.04%

[a] EN_N denotes in-lake N enrichment, DE_N denotes in-lake N depletion, EN_P denotes in-lake P enrichment, DE_P denotes in-lake P depletion.

[b] Internal nutrient cycling intensity ($INCI = \frac{|in{-}lake\ enrichment| + |in{-}lake\ depletion|}{external\ loading}$).

Figure 1. Fluxes and stoichiometry of nutrient cycling across the trophic gradient. Panel (a) and (b) show the N:P ratios of nutrient inflow and outflow, and in-lake enrichment and depletion. Panel (c) shows the N:P distributions of inflow and outflow across the trophic gradient. The dashed line in panel (c) represents the difference between the medians of inflow N:P and outflow N:P. Panel (d) shows the N:P distributions of in-lake enrichment and depletion. The dashed line in panel (d) represents the difference between the medians of in-lake enrichment N:P and in-lake depletion N:P. Panels (e), (f), (g) and (h) show the distributions of fluxes of nutrient inflow, outflow, in-lake enrichment, and in-lake depletion across the trophic gradient (N in panel (e) and (f), P in panel (g) and (h)). The dashed lines in panel(e), (f), (g) and (h) represent the medians of contributions of inflow, in-lake enrichment, outflow, and in-lake depletion to total nutrient input (or output). The white dot in each violin plot represents the median, the thick black line represents the 25th and 75th percentiles, and the thin black line represents the 10th and 90th percentiles.

Figure 2. Global pattern of preferential nutrient retention. Lakes located in Quadrant A (8%) all have a negative net N and P retention. Lakes located in Quadrant B (< 1% of total) retain N but release P, while lakes located in Quadrant D (10%) retain P but release N. Lakes located in Quadrant C (81%) retain both N and P. Lakes located below the black dashed line (87.8%) tend to retain more P than N. EN_N denotes in-lake N enrichment, DE_N denotes in-lake N depletion, EN_P denotes in-lake P enrichment, DE_P denotes in-lake P depletion.

Figure 3. Imbalance of nutrient retention in global lakes. Panels (a) and (b) show the global distribution of in-lake enrichment/depletion ratio in lakes (N in panel (a) and P in panel (b)). Yellow dots represent lakes that positively retain N or P; and the blue dots represent lakes that export N or P to downstream ecosystems.

References

1. Peñuelas, J. & Sardans, J. The global nitrogen-phosphorus imbalance. *Science* **375**, 266–267 (2022).
2. Penuelas, J., Janssens, I. A., Ciais, P., Obersteiner, M. & Sardans, J. Anthropogenic global shifts in biospheric N and P concentrations and ratios and their impacts on biodiversity, ecosystem productivity, food security, and human health. *Global Change Biology* **26**, 1962–1985 (2020).
3. Peñuelas, J. *et al.* Human-induced nitrogen-phosphorus imbalances alter natural and managed ecosystems across the globe. *Nature Communications* **4**, 1–10 (2013).
4. Howarth, R. W., Chan, F., Swaney, D. P., Marino, R. M. & Hayn, M. Role of external inputs of nutrients to aquatic ecosystems in determining prevalence of nitrogen vs. phosphorus limitation of net primary productivity. *Biogeochemistry* **154**, 293–306 (2021).
5. Carnicer, J. *et al.* Global biodiversity, stoichiometry and ecosystem function responses to human-induced C-N-P imbalances. *Journal of Plant Physiology* **172**, 82–91 (2015).
6. Yan, Z. *et al.* Phosphorus accumulates faster than nitrogen globally in freshwater ecosystems under anthropogenic impacts. *Ecology Letters* **19**, 1237–1246 (2016).
7. Elser, J. J. *et al.* Shifts in lake N: P stoichiometry and nutrient limitation driven by atmospheric nitrogen deposition. *Science* **326**, 835–837 (2009).
8. Elser, J. J. *et al.* Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters* **10**, 1135–1142 (2007).
9. Beusen, A. H. W., Bouwman, A. F., van Beek, L. P. H., Mogollón, J. M. & Middelburg, J. J. Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. *Biogeosciences* **13**, 2441–2451 (2016).
10. Qin, B. *et al.* Water depth underpins the relative roles and fates of nitrogen and phosphorus in lakes. *Environmental Science and Technology* **54**, 3191–3198 (2020).
11. Maranger, R., Jones, S. E. & Cotner, J. B. Stoichiometry of carbon, nitrogen, and phosphorus through the freshwater pipe. *Limnology and Oceanography Letters* **3**, 89–101 (2018).
12. Cheng, F. Y. & Basu, N. B. Biogeochemical hotspots: Role of small water bodies in landscape nutrient processing. *Water Resources Research* **53**, 5038–5056 (2017).
13. Grantz, E. M., Haggard, B. E. & Scott, J. T. Stoichiometric imbalance in rates of nitrogen and phosphorus retention, storage, and recycling can perpetuate nitrogen deficiency in highly-productive reservoirs. *Limnology and Oceanography* **59**, 2203–2216 (2014).
14. Paerl, H. W. *et al.* It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science and Technology* **50**, 10805–10813 (2016).
15. Schindler, D. W., Carpenter, S. R., Chapra, S. C., Hecky, R. E. & Orihel, D. M. Reducing phosphorus to curb lake eutrophication is a success. *Environmental Science and Technology* **50**, 8923–8929 (2016).
16. National Lakes Assessment 2012: Technical Report. (2017).
17. Filazzola, A. *et al.* A database of chlorophyll and water chemistry in freshwater lakes. *Scientific Data* **7**, 1–10 (2020).
18. Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications* **7**, 13603 (2016).
19. Carpenter, S. R. Eutrophication of aquatic ecosystems: Bistability and soil phosphorus. *Proc Natl Acad Sci U S A* **102**, 10002–10005 (2005).
20. Søndergaard, M., Bjerring, R. & Jeppesen, E. Persistent internal phosphorus loading during summer in shallow eutrophic lakes. *Hydrobiologia* **710**, 95–107 (2012).
21. Jeppesen, E. *et al.* Lake responses to reduced nutrient loading – an analysis of contemporary long-term data from 35 case studies. *Freshwater Biology* **50**, 1747–1771 (2005).
22. Nowlin, W. H., Evarts, J. L. & Vanni, M. J. Release rates and potential fates of nitrogen and phosphorus from sediments in a eutrophic reservoir. *Freshwater Biology* **50**, 301–322 (2005).
23. Finlay, J. C., Small, G. E. & Sterner, R. W. Human influences on nitrogen removal in lakes. *Science* **342**, 247–250 (2013).

24. Wu, Z., Liu, Y., Liang, Z., Wu, S. & Guo, H. 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 (2017).
25. Harrison, J. A. *et al.* The regional and global significance of nitrogen removal in lakes and reservoirs. *Biogeochemistry* **93**, 143–157 (2008).
26. Seitzinger, S. *et al.* Denitrification across landscapes and waterscapes: a synthesis. *Ecological Applications* **16**, 2064–2090 (2006).
27. Conley, D. J. *et al.* Controlling eutrophication: Nitrogen and phosphorus. *Science* **323**, 1014–1015 (2009).

Materials and Methods

Data

Three datasets were employed in a progressive approach in this study. The first one is NLA2012 (U.S. National Lakes Assessment 2012) dataset for the nutrient budget model development and the estimation of internal nutrient cycling fluxes[1]. Data used in this model include total nitrogen (TN), total phosphorus (TP), chlorophyll-a (Chla), and landscape characteristics from NLA2012, along with lake volumes and discharge rates from HydroLAKES dataset[2]. A total of 596 lakes are identified based on the intersection of the NLA2012 and HydroLAKES databases (Extended Data Fig.1). We use NANI (net anthropogenic nitrogen input) and NAPI (net anthropogenic phosphorus input)[3,4] approach to estimate nutrient inputs of each lake. Briefly, NANI is defined as the sum of five components: atmospheric deposition, fertilizer nitrogen input, agricultural nitrogen fixation, net food and feed import, and non-food crop export. NAPI is the sum of three components: fertilizer phosphorus input, net food and feed import, and non-food crop export[3,4]. We use the data of agricultural census year 2012 to calculate the agricultural nitrogen fixation, net food and feed import, and non-food crop export. Atmospheric deposition is calculated from the data of 2012 in the National Atmospheric Deposition Program[5].

Next, a dataset of 5,622 lakes from Chlorophyll and Water Chemistry database (Extended Data Fig.2) is used for nutrient cycling imbalance verification[6]. The dataset is a combination of the Chlorophyll and Water Chemistry database[6] and HydroLAKES[2], which shares similar distributions on lake characteristics of NLA2012 dataset (Extended Data Fig.3). A generalized linear model (GLM) is trained for the 596 lakes in NLA2012, to examine the relationship between the modeled internal nutrient cycling fluxes and a series of physical characteristics of lakes. Data used in the GLM training include trophic state and surface water temperature from NLA2012 and water residence time, lake area, and lake depth from HydroLAKES. Then the fitted GLM is applied to predict the patterns of internal nutrient cycling fluxes for the 5,622 lakes.

A dataset of 1.4 million lakes from HydroLAKES database (including the above dataset of 5,622 lakes) is used for global upscaling of retention flux and preferential retention estimation. The two most significant variables, lake surface area and trophic state, are used in the estimation. A two-layer Monte Carlo approach is adopted to predict the global fluxes from the distribution of internal fluxes in the results of nutrient budget model. The joint distribution of global lake areas and trophic states are generated from the Chlorophyll and Water Chemistry database[6] and HydroLAKES[2].

Nutrient Budget Model

Model Development

A nutrient budget model is developed to describe the fluxes and internal nutrient cycling processes. The internal nutrient cycling is defined with two processes, including (a) nutrient in-lake enrichment, which represents the internal input of nutrients such as sediment release; and (b) nutrient in-lake depletion, which represents the internal removal of nutrients such as sedimentation and denitrification. Previous studies have shown that the intensity of in-lake enrichment is directly related to the abundance of algae in lakes, as most often indicated by Chla[7,8]. Lakes with high algal biomass tend to have high nutrient enrichment flux. In-lake nutrient enrichment is described as a modified Michaelis-Menten function of Chla, similar to Carpenter (2005)[9]. In-lake nutrient depletion is described as a first order process of TN (or TP). We followed the equations in Sinha et al.[10,11] to estimate the nutrient loading of each lake from NANI and NAPI (Eq.3 and 4), using the same hierarchical structure as below.

$$\frac{dTN}{dt} = \frac{L_N}{V} + \frac{a_N}{V} \cdot \frac{Chla^{b_N}}{Chla^{b_N} + m_N^{b_N}} - S_N \cdot TN - \frac{F_{out}}{V} \cdot TN \quad (1)$$

$$\frac{dTP}{dt} = \frac{L_P}{V} + \frac{a_P}{V} \cdot \frac{Chla^{b_P}}{Chla^{b_P} + m_P^{b_P}} - S_P \cdot TP - \frac{F_{out}}{V} \cdot TP \quad (2)$$

$$ln(L_N) = \theta_1^N * \mathrm{asinh}(NANI/2) + \theta_2^N * LU_{forest} + \theta_3^N * LU_{wetland} + \theta_4^N \quad (3)$$

$$ln(L_P) = \theta_1^P * \mathrm{asinh}(NAPI/2) + \theta_2^P * LU_{forest} + \theta_3^P * LU_{wetland} + \theta_4^P \quad (4)$$

where L_N is the external loading input of TN ($g\ N \cdot day^{-1}$); L_P is the external loading input of TP ($g\ P \cdot day^{-1}$); a_N is the coefficient of in-lake N enrichment ($g\ N \cdot day^{-1}$); m_N is the half saturation constant of Chla for in-lake N enrichment ($mg\ Chl \cdot m^{-3}$); b_N is the shape parameter of Chla for in-lake N enrichment; a_P is the coefficient of in-lake P enrichment ($g\ P \cdot day^{-1}$); m_P is the half saturation constant of Chla for in-lake P enrichment ($mg\ Chl \cdot m^{-3}$); b_P is the shape parameter of Chla for in-lake P enrichment; S_N is the rate of in-lake N depletion (day^{-1}); S_P is the rate of in-lake P depletion (day^{-1}); F_{out} is the discharge rate ($m^3 \cdot day^{-1}$); V is the lake volume (m^3); and asinh is the inverse hyperbolic sine function. The θ_1^N, θ_2^N, θ_3^N, and θ_4^N are the parameters to estimate the external loading of TN. The θ_1^P, θ_2^P, θ_3^P, and θ_4^P are the parameters to estimate the external loading of TP. The LU_{forest} is the percentage of the basin area classified as forest (%); and $LU_{wetland}$ is the percentage of the basin area classified as wetland (%).

The in-lake N enrichment (EN_N) and in-lake P enrichment (EN_P) of a lake are calculated as Eq.5 and Eq.6, i.e., in-lake N depletion (DE_N) and in-lake P depletion (DE_P) of a lake are calculated as Eq.7 and Eq.8.

$$EN_N = \frac{a_N}{V} \cdot \frac{Chla^{b_N}}{Chla^{b_N} + m_N^{b_N}} \quad (5)$$

$$EN_P = \frac{a_P}{V} \cdot \frac{Chla^{b_P}}{Chla^{b_P} + m_P^{b_P}} \quad (6)$$

$$DE_N = S_N \cdot TN \quad (7)$$

$$DE_P = S_P \cdot TP \quad (8)$$

Bayesian Inference

We adopt a Bayesian hierarchical framework to implement and analyze the nutrient budget model of the NLA2012 dataset. Due to the heterogeneity in the attributes of both water body and basin of lakes, a hierarchical structure is applied to the equations of mass balance and loading estimation which allowed for variability in model parameters at the level of ecological region. The lakes were categorized into 8 regions based on the Level I definition of ecological regions of North America[12]. The lakes within the same ecological region share similar features of landscape, soil characteristics, and climate, which lead to similar patterns of nutrient flux producing and nutrient cycling behaviors. Thus, the lakes located in the same ecological region use the same set of parameters. More specifically, parameters θ_1^N, θ_2^N, θ_3^N, θ_4^N, θ_1^P, θ_2^P, θ_3^P, θ_4^P, a_N, a_P, b_N, b_P, S_N, and S_P, denoted by p, each follow a normal distribution:

$$p_i \sim normal(\mu_p, \sigma_p) \quad (9)$$

where p_i is the parameter p that used in ecoregion i. It is drawn from a normal distribution with μ_p as mean and σ_p as standard deviation (Eq.9). Parameter inference is carried out in the software Stan (version 2.28) with Hamiltonian Monte Carlo (HMC) algorithm, which has been shown to have superior speed and performance for fitting complex dynamic models compared to other Markov Chain Monte Carlo (MCMC) methods[13]. We use No-U-Turn Sampler (NUTS) in Stan to avoid manual selection of application-specific tuning parameters. Four HMC chains were run for

2000 iterations (including 1000 iterations for warm-up). The $\hat{R}$ convergence diagnostic is monitored for model fits to ensure $\hat{R}$<1.05. The prior distributions of parameters are shown in Extended Data Table 1. The posterior distributions of the parameters are shown in Extended Data Fig.4. The hierarchical parameters are shown in Fig.S1-S2 and model performance is shown in Fig.S3 and Fig.S4.

Generalized Linear Model

The relationship between nutrient retention ability (f) and physical characteristics was estimated for the lakes in NLA2012 dataset using a generalized linear model (GLM), i.e.

$$\begin{gathered} ln(f) = \beta_0 + \beta_1 ln(WRT) + \beta_2 ln(SWT) + \beta_3 ln(Depth) + \\ \beta_4 ln(Area) + \beta_5 TI_2 + \beta_6 TI_3 + \beta_7 TI_4 \end{gathered} \tag{10}$$

where f is the estimated dependent variable representing either E_N/L_N, E_P/L_P, D_N/L_N, or D_P/L_P; β_0, β_1, β_2, β_3, β_4, β_5, β_6, and β_7 are parameters of the GLM; WRT is water residence time (days); SWT is the water temperature 0.5 meter below surface ($^\circ C$); Depth is the lake average depth (meters); TI_2, TI_3, and TI_4 are ordinal dummy variables representing trophic states, parameterized relative to trophic state level 1 (the oligotrophic state). Four trophic states are grouped by Chla concentration, oligotrophic (<2 μg/L), mesotrophic (2-7 μg/L), eutrophic (7-30 μg/L), and hyper-eutrophic (>30 μg/L). All the continuous variables used in the GLM are log-transformed because the distributions of these variables approximately follow a log-normal distribution. Instead of predicting in-lake nutrient enrichment and depletion directly, the ratios of in-lake nutrient enrichment (or depletion) to external nutrient loading are predicted to normalize the external impact across different nutrient inputs to the lakes. We use the *glm* function in R (version 4.1.2) to fit the model. The model performance is shown in Fig.S5 and S6 and the results are shown in Extended Data Table 2-5. The derived GLM results of NLA2012 were then used to verify the imbalance of nutrient cycling for a lager dataset of 5,622 lakes with similar distributions of the dependent variables (Extended Data Fig.2 and Extended Data Fig.3).

Up-scaling Global Estimation

Global lakes can be divided into different categories according to lake size (area), trophic state (Chla), and other characteristics. In order to upscale the internal nutrient cycling fluxes, it is necessary to know the global joint distribution of these characteristics. Lake surface area and trophic state, two significant variables with available data globally, are used in the estimation. A global joint distribution of lake surface area and trophic state is generated from the Chlorophyll and Water Chemistry database[6] and HydroLAKES[2] (Table S1). More specifically, global lakes are grouped into 9 categories of surface area ranging from 0.001 km^2 to 1×10^6 km^2. In each category of lake surface area, lakes are then grouped into 4 categories based on trophic states. Since the data are skewed, we perform a logarithmic transformation before Bayesian inference. We then apply a Monte Carlo approach to predict the areal internal nutrient cycling fluxes (EN_N, EN_P, DE_N, DE_P) for each unique combination of lake size and trophic state from the distribution of the nutrient budget model results. Next, we multiplied the predicted areal fluxes for each joint category by the total surface area of the category, then summed across the categories of trophic state to find the total fluxes of each size category. The global internal nutrient cycling fluxes are generated by summing across the categories of lake size. The results of different trophic states are shown in Table S2. We further grouped the global 1.4 million lakes in the HydroLAKES dataset into the above categories and roughly estimate the internal cycling fluxes for each lake plotted in Extended Data Figure 5.

Data availability

The data of National Lake Assessment 2012 were obtained from USEPA (https://www.epa.gov/national-aquatic-resource-surveys/nla). The Chlorophyll and Water Chemistry database were retrieved from Scientific Data (https://doi.org/10.1038/s41597-020-00648-2). The HydroLAKES dataset was retrieved from Global HydroLAB (https://wp.geog.mcgill.ca/hydrolab/hydrolakes/). The processed data to reproduce the results in this study are available at GitHub (https://github.com/zhenwu0728/Preferential_Nutrient_Rentention_in_Global_Lakes, DOI: 10.5281/zenodo.5944260).

Code availability

The CmdStan (version 2.28) software used for the nutrient budget model is available at https://mc-stan.org/users/interfaces/cmdstan. Julia (version 1.6.5) used as the interface to run CmdStan is available at https://julialang.org. R (version 4.1.2) used for GLM analysis is available from the R Core Team (https://www.r-project.org/). Python (version 3.10) used for global up-scaling analysis is available at https://www.python.org/. The codes to reproduce the results in this study is available at https://github.com/zhenwu0728/Preferential_Nutrient_Rentention_in_Global_Lakes.

References

1. National Lakes Assessment 2012: Technical Report. (2017).
2. Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications* **7**, 13603 (2016).
3. Howarth, R. W. *et al.* Riverine inputs of nitrogen to the North Atlantic Ocean: fluxes and human influences. *Biogeochemistry* **35**, 75–139 (1996).
4. Hong, B. *et al.* Advances in NANI and NAPI accounting for the Baltic drainage basin: spatial and temporal trends and relationships to watershed TN and TP fluxes. *Biogeochemistry* **133**, 245–261 (2017).
5. National Atmospheric Deposition Progra. Total Deposition Maps, v2018.01.
6. Filazzola, A. *et al.* A database of chlorophyll and water chemistry in freshwater lakes. *Scientific Data* **7**, 1–10 (2020).
7. Søndergaard, M., Bjerring, R. & Jeppesen, E. Persistent internal phosphorus loading during summer in shallow eutrophic lakes. *Hydrobiologia* **710**, 95–107 (2012).
8. Jeppesen, E. *et al.* Lake responses to reduced nutrient loading – an analysis of contemporary long-term data from 35 case studies. *Freshwater Biology* **50**, 1747–1771 (2005).
9. Carpenter, S. R. Eutrophication of aquatic ecosystems: Bistability and soil phosphorus. *Proc Natl Acad Sci U S A* **102**, 10002–10005 (2005).
10. Sinha, E. & Michalak, A. M. Precipitation dominates interannual variability of riverine nitrogen loading across the continental United States. *Environmental Science and Technology* **50**, 12874–12884 (2016).
11. Sinha, E., Michalak, A. M. & Balaji, V. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science (1979)* **357**, (2017).
12. Commission for Environmental Cooperation. Ecological regions of North America Toward a Common Perspective. (1997).
13. Monnahan, C. C., Thorson, J. T. & Branch, T. A. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution* **8**, 339–348 (2017).