

The relative role of nutrient inputs, climate, land use, and spatial patterns on chlorophyll *a* concentrations worldwide

Arnab Shuvo

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTERS OF SCIENCE

Graduate Program in
BIOLOGY

York University
Toronto, Ontario

August 2020

© Arnab Shuvo, 2020

ABSTRACT

Water quality degradation is one of the largest threats to freshwater ecosystems, and understanding how multiple drivers such as nutrients, climate, morphology, and land use impact water quality is of high interest because it will help mitigate the potential consequences of degraded waters. We used chlorophyll *a* as a proxy to primary production to quantify: 1) the relative importance of drivers on water quality in freshwater lakes globally, and 2) identify the spatial patterns and signatures of chlorophyll and its drivers in freshwater lakes across the US. We found that total phosphorus (TP) inputs at the local spatial scale was the single most influential driver of chlorophyll levels in freshwater lakes globally and in the US. Climate, including warmer temperatures, increased solar radiation inputs, and wetter conditions, further contributed to elevated chlorophyll concentrations. As such, we suggest management efforts should continue to focus on drivers acting at local spatial scales, in particular TP, to mitigate primary production in lakes worldwide.

ACKNOWLEDGMENTS

I would like to thank my supervisor Dr. Sapna Sharma for her dedicated support and guidance throughout my graduate and undergraduate degree. As a first-generation grad student there was a whole lot of unknowns and grad school was often intimidating and confusing, however Dr. Sharma continuously provided encouragement and was always willing and enthusiastic to assist in any way she could. Her invaluable insights, attention to detail, and constructive feedback has led this research project to all its success, and I am truly grateful to have had her as mentor throughout my academic career. Without any doubts I can say, I had the best supervisor ever!

I would also like to thank Dr. Alex Filazzola and Dr. Kevin Blagrove, to whom I have asked numerous stats and coding related questions. However, they were always patient and willing to help. I would also like to extend a large thank you to Dr. Usman Khan, Dr. Lewis Molot, Dr. Catherine O'Reilly, Dr. Derek Gray, Dr. Roberto Quinlan, and Dr. Dawn Bazely. Dr. Khan and Dr. Molot provided excellent feedback on my manuscripts and challenged me by asking thought provoking questions during our meetings. I have also had the pleasure of co-authoring on research papers with Dr. O'Reilly, Dr. Gray, and Dr. Quinlan who have all helped improve my writing skills as well as my foundational knowledge of complex limnological concepts. Dr. Bazely has always been extremely supportive (since undergrad) and has continued to be throughout my grad career by constantly checking in on academic progress and my general well being.

Prior to the pandemic when we could go into the lab for work, it was always great collaborating with Octavia, Luke, Carrie, and Tanzil. The teamwork was always strong, and the results were great. I thank you all for being so wonderful and supportive over the past two years!

Outside of the lab, I would like to acknowledge all the contributors for their wonderful collaboration and sharing their resources with me. I am also thankful to all my friends who supported me throughout this entire process, believed in me in the moments when I had doubts, and even pulled some all nighters with me.

Finally, I would like to extend my deepest gratitude to my family for providing me with unfailing support and the resources which have helped me get this far. This accomplishment would not have been possible without them. Thank you!

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	vi
LIST OF SUPPLEMENTARY MATERIAL	viii
GENERAL INTRODUCTION.....	ix
CITATIONS FOR M.Sc. THESIS:	xiv
REFERENCES	xv
Chapter 1: Total phosphorus and climate are equally important predictors of water quality in lakes: A global analysis	xviii
MY CONTRIBUTION	xix
SUMMARY	xx
INTRODUCTION	1
METHODS	5
2.1 Data Acquisition and Dataset Compilation.....	5
2.2 Data analyses	6
2.3 Drivers of Water Quality	7
RESULTS	9
3.1 Drivers of Water Quality including Solar Radiation (using D+SR)	9
3.2 Drivers of Water Quality excluding Solar Radiation.....	10
DISCUSSION	12
CONCLUSION.....	17
ACKNOWLEDGMENTS	18
REFERENCES	19
FIGURES.....	24
SUPPLEMENTARY MATERIALS	30
Chapter 2: Local, regional, and watershed spatial patterns structure chlorophyll <i>a</i> concentrations in lakes across the contiguous United States	34
MY CONTRIBUTION	35
SUMMARY	36
INTRODUCTION	37
METHODS	40
2.1 Data Acquisition and Dataset Compilation.....	40

2.2 Data analyses	42
2.3 Spatial patterns and drivers of water quality.....	43
RESULTS	45
Spatial patterns and scales of water quality	45
Spatial signatures of chlorophyll a and its drivers	45
Relative importance of drivers of water quality	46
DISCUSSION	48
Spatial patterns and scales of water quality	49
Spatial signatures of chlorophyll a and its drivers	50
Relative importance of drivers of water quality	51
ACKNOWLEDGEMENTS	54
REFERENCES	55
FIGURES	60
SUPPLEMENTARY MATERIALS	64
GENERAL CONCLUSION	66
REFERENCES	70

LIST OF FIGURES

Chapter 1:

Figure 1. Maps showing lake locations for which water chemistry and climate data were collected for (a) 2561 lakes for which solar radiation data were available in addition to water chemistry and climate data (water temperature, precipitation, and cloud cover), (b) 3878 lakes for which there were only water chemistry and climate data available.

Figure 2. Stacked bar plot representing the proportion of variation explained by each variable in the random forest model. These variables are all drivers of chlorophyll in (A) the 2561-lake dataset which include solar radiation (D+SR) and (B) the 3878-lake dataset which does not include solar radiation data (D-SR). Variables are separated in groups characterized by nutrients (total phosphorus), climate (spring and summer solar radiation, cloud cover, precipitation, and air temperature), and morphology (elevation, lake area, mean depth, residence time, volume, watershed area).

Figure 3. Chlorophyll concentrations plotted against the most influential climate and geomorphometric drivers for the 2561-lake dataset that includes solar radiation (D+SR), with linear models for (a) spring solar radiation ($R^2_{adj} = 0.041$), (b) total phosphorus ($R^2_{adj} = 0.47$), (c) summer air temperature ($R^2_{adj} = 0.14$), and (d) summer cloud cover ($R^2_{adj} = 0.045$).

Figure 4. Chlorophyll concentrations plotted against the most influential drivers for the 3878-lake dataset without solar radiation (D-SR), with linear models for (a) total phosphorus concentrations ($R^2_{adj} = 0.0094$), (b) summer air temperature ($R^2_{adj} = 0.13$), (c) summer cloud cover ($R^2_{adj} = 0.023$) and (d), mean lake depth ($R^2_{adj} = 0.062$).

Figure 5. Principal Component Analysis for (a) the 2561-lake dataset (D+SR) and (b) the 3878-lake dataset (D-SR). The most variation is explained by Principal Component 1 (Dim1), followed by the second principal component axis (Dim2). The arrows depict the environmental variables considered in the analysis. The longer the arrow, the more influential the variable. The angle between arrows suggests the correlation between variables. The acronyms are as follows: TP (total phosphorus), Temp (temperature), Ppt (precipitation), srad (solar radiation), Cldcvr (cloud cover)

Chapter 2:

Figure 1. Maps showing lake locations for which water chemistry and climate data were collected in the United States. Lake trophic status is indicated by colour based on the Carlson trophic index Chl a thresholds, where measures below 0.0026 mg/L are considered oligotrophic, measures between 0.0026 and 0.020 mg/L are mesotrophic, between 0.020 and 0.056 mg/L are considered eutrophic, and above 0.056 are hyper-eutrophic.

Figure 2. Plots of Moran's I showing spatial autocorrelation in chlorophyll, phosphorus values, percentage cropland, mean depth, elevation, pH, summer precipitation, summer solar radiation, and summer temperature values ranging from 0-4000kms.

Figure 3. Bubble plots highlighting the four MEM's that explained the highest variation in Chl a. The combination of all significant MEM variables explained 36% of the variation in Chl a. MEM 73 (a) and MEM25 (c) explain 3.4% of the variation and reflect spatial patterns acting at intermediate and watershed level scales. MEM364 (b) and MEM378 (d) explain 2.8% of the variation and represent fine spatial patterns acting at local spatial scales.

Figure 4. Variation partitioning analysis representing the four major groups and their relative influence on chlorophyll production in freshwater lakes across the United States. X1 is representative of local variables: TP, lake area, volume, average depth, residence time, elevation, watershed area. X2 represents climate: summer temperature, spring temperature, spring solar radiation, summer solar radiation, summer precipitation, summer cloud cover, and spring cloud cover. X3 represents the land use variables and human population: croplands, woody savannas, deciduous broadleaf forests, mixed forests, and population. X4 highlights the spatial structure through the 88 significant MEMs.

LIST OF SUPPLEMENTARY MATERIAL

Chapter 1:

Table S1. A summary table of the geomorphometric, water chemistry, and climate data extracted for both datasets encompassing 2561 (data including solar radiation data; (D+SR)) and 3878 lakes (dataset excluding solar radiation data (D-SR)). The units for each variable are listed next to the variable of interest. The mean and range has been calculated for each variable. *Data sourced from HydroLAKES (Messenger et al. 2016). Shoreline development measures the degree of deviation of a lake's surface shape from a circle; a value of 1 implies a perfectly circular shape and higher values indicate increased shoreline sinuosity (Messenger et al. 2016).

Figure S1. Kendall's Tau correlations between chlorophyll *a*, total phosphorus (TP), summer climate variables, spring climate variables, and geomorphometric characteristics for a) the dataset with solar radiation (D+SR). Warm colours indicate positive relationships, whereas cooler colours represent negative relationships. Squares which are left blank and white are done so because they are not significantly related ($P > 0.05$). The acronyms are as follows: TP (total phosphorus), Temp (temperature), Ppt (precipitation), srad (solar radiation), Clcdvr (cloud cover).

Figure S2. Kendall's Tau correlations between chlorophyll *a*, total phosphorus (TP), summer climate variables, spring climate variables, and geomorphometric characteristics for the larger dataset without solar radiation (D-SR). Warm colours indicate positive relationships, whereas cooler colours represent negative relationships. Squares which are left blank and white are done so because they are not significantly related ($P > 0.05$). The acronyms are as follows: TP (total phosphorus), Temp (temperature), Ppt (precipitation), srad (solar radiation), Clcdvr (cloud cover).

Chapter 2:

Figure S1. Random forest model (2240 lakes dataset) assessing the key drivers of chlorophyll in lakes across the United states. The drivers assessed were TP, climate, lake morphology, human population, and land use. The random forest model explained 61% variation and the mean of squared residuals was 0.1109.

Figure S2. Correlation plot (2240 lakes dataset) highlighting the relationship between chlorophyll (response variable) and total phosphorus (TP), climate, land use, morphology, and human population (predictor variables). Relationships between predictors variables are also shown. Values closer to +1 signify a strong positive relation, whereas -1 suggests a strong negative relationship.

GENERAL INTRODUCTION

Globally there are over 118 million lakes (Verpoorter et al., 2014; Woolway et al., 2020) that contain 87% of Earth's liquid freshwater (Gleick, 1993; Woolway et al., 2020) and these freshwaters are earth's most valuable resource (Jenny et al., 2020). Lakes provide key ecosystem services such as provisioning (food, drinking water, industrial water), regulating (water flow regulations, local climate controls), supporting (habitats for reproduction for flora and fauna, support for migratory species, biodiversity hotspots), and cultural services to humans (aesthetics, spiritual experience, cognitive development). However, the continuous growth of the global human population has extensively altered the planet's freshwaters (Carpenter et al., 2011) and anthropogenic stressors such as changes in land use and land cover (agriculture and urbanization) along with climate change are changing the physical, chemical, and biological make up of aquatic ecosystems (Carpenter et al., 2011; Michalak, 2016). Through increases in agricultural practices and urbanized regions, local inputs of nutrients such as phosphorus are more readily found in lake basins and watersheds. Coupled with the effects of climate change such as increased water temperatures (Collins et al., 2019) and precipitation levels (Sinha et al., 2017), greater nutrient loading is occurring in lakes which is resulting in more frequent algal blooms and algal biomasses. Such degradation of water quality results in the loss of ecosystem services that are beneficial to humans and other species alike (Carpenter et al., 2011; Michalak, 2016). Further, recent studies suggests lakes that have been subject to ongoing degradation are more susceptible to new threats from the synergies between multiple stressors and may be less resistant to eutrophication (Jenny et al., 2020). Understanding how multiple drivers such as nutrients, climate, and land use impact water quality is of high interest and will help develop

effective environmental management plans and mitigate the potential consequences of degraded waters (Vaughan et al., 2019).

Nutrient concentrations, particularly total phosphorus, are well established as the key limiting factor for primary production and chlorophyll (Schindler, 1977; McCauley et al., 1989; Abell et al., 2012; Quinlan et al., in review). The relationship between total phosphorus (TP) and chlorophyll allows TP to be used when determining lake trophic status (Fee, 1979). As such, managing TP concentrations is a common step in lake eutrophication management because without TP reduction, there are larger algal bloom events occurring (Ho et al., 2019). The importance of TP on primary production in lakes was studied by early limnologists such as Hutchinson and Nauman, but it was Schindler who showed its importance through a multi-year experiment at the experimental lake area (ELA) which showed it was indeed TP that was the limiting reagent and not nitrogen (Schindler, 1977). Schindler (1977) also suggests that lakes can compensate for low carbon and nitrogen levels by tapping into the atmospheric reservoir but that same is not possible for phosphorus. In addition, studies have suggested in oligotrophic lakes (Carpenter et al., 1992; Lampert & Sommer, 1997; Wetzel, 2001) phosphorus is indefinitely the limiting reagent, however even in other aquatic systems (mesotrophic and eutrophic systems) (Horne & Goldman, 1994; Dodds, 2002; Bronmark & Hansson, 2005) phosphorus is also quite closely linked to primary production. However, owing to the influence of climate change, recent studies have challenged these paradigms for both lakes experiencing extreme eutrophication (Paerl et al., 2019) and oligotrophication (Huser et al., 2018). Quinlan et al. (in review) has shown that as range in phosphorus increases (low-high TP values) at broad spatial scales, TP's relationship with chlorophyll becomes sigmoidal, wherein there are lags at

the beginning of the sigmoid and other drivers such as light may be more influential near the end where TP levels have been saturated. These ranges at broad scales are being greatly influenced by climate change. Different regions globally are experiencing never before experienced climactic conditions, resulting the ranges are becoming wider.

Freshwater lakes are directly affected by climate change (Carpenter et al., 2011; Woolway et al., 2020) through increased air temperatures (Carpenter et al., 2011; O'Reilly et al., 2015), solar radiation (Sanchez-Lorenzo et al., 2015; Wild et al., 2015), and precipitation (Carpenter et al., 2011; O'Reilly et al., 2015). Warming of air temperatures has been strongly linked to warming of water temperatures over several decades (Adrian et al., 2009; O'Reilly et al., 2015; Sharma et al., 2015; Richardson et al., 2017), however patterns of geographic heterogeneity between lakes and surface water warming suggests other local or regional drivers may be influencing this variability (Schneider & Hook, 2010; Kraemer et al., 2015; O'Reilly et al., 2015; Richardson et al., 2017). Globally over 90% of observed lakes are exhibiting warming trends (Schneider & Hook, 2010; O'Reilly et al., 2015; Sharma et al., 2015; Woolway et al., 2016; Richardson et al., 2017; Bachman et al., 2019) and such increased temperatures may produce the excessive proliferation of phytoplankton species (Paerl et al. 2001; O'Reilly et al. 2015) and harmful algal blooms (Kraemer et al., 2015; Bachman et al., 2019). Solar radiation is closely linked to warming air and water temperatures. Currently the earth is undergoing a period of global brightening (Wild et al., 2015) and the increased light availability can be attributed to larger algal bloom. Further, the increased radiation that is heating up air and water temperatures reduces the ice season duration, which allows for longer growing seasons for algae (Carpenter et al., 2011; Woolway et al., 2020). Large precipitation events exacerbate the issue to an even greater degree

and degrade water quality through the addition of nutrients from runoff, surface waters, and stormwaters (Sinha et al., 2017). In addition, large storms erode riverbanks and other terrestrial sediment sources which end up in the lake and contribute greatly to the lake's annual N and P budget (Reichwaldt & Ghadouani, 2012). Collins et al. (2019) suggests the combined effects of warmer temperatures and precipitation can be attributed to increased algal biomasses.

Lakes that are close to agricultural lands (Wagner et al., 2011) or urbanized regions (Hopkins et al., 2015) receive large nutrient inputs, particularly phosphorus, from runoff and surface waters after precipitation events. These increased phosphorus inputs degrade water quality due to the proliferation of algal blooms (Jeppesen et al., 2009) and decreases in water clarity (Rose et al., 2017). Agricultural land is comprised of croplands and pasturelands both of which contribute phosphorus to runoff which ends up in lakes. Croplands extensively use pesticides and fertilizers rich in phosphorus (Hu et al., 2013), while large animals such as cattle, and sheep contribute to the phosphorus inputs through soil erosions and manure (Cournane et al., 2011). Aside from agricultural practices, urbanization is happening quite rapidly at the global level and is increasing the amount of impervious surfaces present (Hopkins et al., 2015). Runoff from these impervious surfaces are unable to percolate into the ground due to the lack of soils, and thus directly ends up in lakes where the addition of nutrients increases the production of algae (Wurtsbaugh et al., 2019).

My thesis aims to examine the effects of nutrients, climate change, and land use on water quality, specifically chlorophyll levels in freshwater lakes. I explore these effects at two spatial scales in two chapters: first by focusing on 3878 freshwater lakes in 41 countries, and second by

employing a finer examination of spatial patterns in water quality across the United State in 2240 lakes.

In the first chapter, we examined water quality trends between 3878 lakes at the global scale with lakes available from every continent. This is the first study that has attempted to quantify water quality patterns at a global scale while incorporating physical, chemical, and biological drivers. In this study we asked: what is the relative importance of nutrient inputs, climate conditions, and lake morphology on chlorophyll *a* levels in lakes globally? We hypothesized that phosphorus would be the most influential driver of chlorophyll *a* because of its multiple sources of input, but climate would also be highly influential because of its broad spatial scale and its many drivers such as temperature, solar radiation, and precipitation. In the second chapter, we focused on 2240 lakes spatially distributed across the United States. With this US dataset, we aimed to quantify the spatial patterns in water quality and determine the relevant spatial signatures of chlorophyll, as well as phosphorus, climate, land use and morphological variables. Little is known about spatial structures beyond local scales, so we attempted to understand spatial structures at the local scale but also at the watershed and regional scale across the US. We hypothesized that chlorophyll *a* levels and trophic status will show spatial structure in response to spatially structured environmental variables and that after factoring in the major spatially structured environmental variables, space will no longer be important.

CITATIONS FOR M.Sc. THESIS:

Over the course of the MSc degree, I have submitted four papers as a co-author. The second chapter will be submitted to *Water*.

Shuvo, A., Gray, D., Moslenko, L., Blaggrave, K., Ewins, C., Filazzola, A., Mahdiyan, O., O'Reilly, CM., Quinlan, R., Sharma, S. (Submitted). Local, regional, and watershed spatial patterns structure chlorophyll *a* concentrations in lakes across the contiguous United States. *Water*.

Shuvo, A., O'Reilly, CM., Blaggrave, K., Ewins, C., Filazzola, A., Gray, D., Mahdiyan, O., Moslenko, L., Quinlan, R., Sharma, S. (In review). Total phosphorus and climate are equally important predictors of water quality in lakes: A global analysis. *Aquatic Sciences*.

Filazzola, A., Mahdiyan O., **Shuvo, A.**, Ewins, C., Moslenko, L., Sadid, T., Blaggrave, K., Gray, D., Quinlan, R., O'Reilly, CM., Sharma, S. (In review). A global database of chlorophyll and water chemistry in freshwater lakes. *Scientific Data*.

Quinlan, R., A. Filazzola, A., Mahdiyan, O., **Shuvo, A.**, Blaggrave, K., Ewins, C., Moslenko, L., Gray, D., O'Reilly, CM., Sharma, S. (In review). Relationships of total phosphorus and chlorophyll in lake worldwide. *Limnology and Oceanography*

Vucic, J.M., Cohen, R.S., Gray, D.K., Murdoch, A.D., **Shuvo, A.**, Sharma, S. (2018). Young gravel pit lakes along Canada's Dempster Highway: How do they compare with natural lakes? *Arctic, Antarctic, and Alpine Research* 51: 25-39.

REFERENCES

- Abell, J. M., Özkundakci, D., Hamilton, D. P., & Jones, J. R. (2012). Latitudinal variation in nutrient stoichiometry and chlorophyll-nutrient relationships in lakes: a global study. *Fundamental and Applied Limnology/Archiv für Hydrobiologie*, 181(1), 1-14.
- Adrian, R., O'Reilly, C.M., Zagarese, H., Baines, S.B., Hessen, D.O., Keller, W., Livingstone, D.M., Sommaruga, R., Straile, D., Van Donk, E. and Weyhenmeyer, G.A. (2009). Lakes as sentinels of climate change. *Limnology and oceanography*, 54(6part2), 2283-2297.
- Bachmann, R. W., Sharma, S., Canfield, D. E., & Lecours, V. (2019). The Distribution and Prediction of Summer Near-Surface Water Temperatures in Lakes of the Coterminous United States and Southern Canada. *Geosciences*, 9(7), 296.
- Brönmark, C., & Hansson, L. A. (2017). *The biology of lakes and ponds*. Oxford University Press.
- Carpenter, S. R., Cottingham, K. L., & Schindler, D. E. (1992). Biotic feedbacks in lake phosphorus cycles. *Trends in Ecology & Evolution*, 7(10), 332-336.
- Carpenter, S. R., Stanley, E. H., & Vander Zanden, M. J. (2011). State of the world's freshwater ecosystems: physical, chemical, and biological changes. *Annual review of Environment and Resources*, 36, 75-99.
- Collins, S.M., Yuan, S., Tan, P.N., Oliver, S.K., Lapierre, J.F., Cheruvilil, K.S., Fergus, C.E., Skaff, N.K., Stachelek, J., Wagner, T. and Soranno, P.A. (2019). Winter precipitation and summer temperature predict lake water quality at macroscales. *Water Resources Research*, 55(4), 2708-2721.
- Cournane, F. C., McDowell, R., Littlejohn, R., & Condon, L. (2011). Effects of cattle, sheep and deer grazing on soil physical quality and losses of phosphorus and suspended sediment losses in surface runoff. *Agriculture, ecosystems & environment*, 140(1-2), 264-272.
- Dodds, W. (2002). *Freshwater ecology: concepts and environmental applications*. Elsevier.
- Fee, E. J. (1979). A relation between lake morphometry and primary productivity and its use in interpreting whole-lake eutrophication experiments. *Limnology and Oceanography*, 24(3), 401-416.
- Gleick, P. H. (1993). Water and conflict: Fresh water resources and international security. *International security*, 18(1), 79-112.
- 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.
- Hopkins, K.G., Morse, N.B., Bain, D.J., Bettez, N.D., Grimm, N.B., Morse, J.L., Palta, M.M., Shuster, W.D., Bratt, A.R. and Suchy, A.K. (2015). Assessment of regional variation in streamflow responses to urbanization and the persistence of physiography. *Environmental science & technology*, 49(5), 2724-2732.
- Horne, A. J., & Goldman, C. R. (1994). *Limnology*. McGraw-Hill.
- Hu, X., Roberts, D.P., Xie, L., Maul, J.E., Yu, C., Li, Y., Zhang, S. and Liao, X. (2013). Development of a biologically based fertilizer, incorporating *Bacillus megaterium* A6, for improved phosphorus nutrition of oilseed rape. *Canadian journal of microbiology*, 59(4), 231-236.
- Huser, B. J., Futter, M. N., Wang, R., & Fölster, J. (2018). Persistent and widespread long-term phosphorus declines in Boreal lakes in Sweden. *Science of the total environment*, 613, 240-249.

- Jenny, J.P., Anneville, O., Arnaud, F., Baulaz, Y., Bouffard, D., Domaizon, I., Bocaniov, S.A., Chèvre, N., Dittrich, M., Dorioz, J.M. and Dunlop, E.S. (2020). Scientists' Warning to Humanity: Rapid degradation of the world's large lakes. *Journal of Great Lakes Research*.
- Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K.M., Andersen, H.E., Lauridsen, T.L., Liboriussen, L., Beklioglu, M., Özen, A. and Olesen, J.E. (2009). Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. *Journal of environmental quality*, 38(5), 1930-1941.
- Kraemer, B.M., Anneville, O., Chandra, S., Dix, M., Kuusisto, E., Livingstone, D.M., Rimmer, A., Schladow, S.G., Silow, E., Sitoki, L.M. and Tamatamah, R. (2015). Morphometry and average temperature affect lake stratification responses to climate change. *Geophysical Research Letters*, 42(12), 4981-4988.
- McCauley, E., Downing, J.A., & Watson, S. (1989). Sigmoid relationships between nutrients and chlorophyll among lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 46 (7), 1171-1175.
- Michalak, A. M. (2016). Study role of climate change in extreme threats to water quality. *Nature*, 535(7612), 349-350.
- O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J., Schneider, P., Lenters, J.D., McIntyre, P.B., Kraemer, B.M. and Weyhenmeyer, G.A. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, 42(24), 10-773.
- Paerl, H. W., Fulton, R. S., Moisander, P. H., & Dyble, J. (2001). Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *TheScientificWorldJournal*, 1.
- Paerl, H.W., Havens, K.E., Xu, H., Zhu, G., McCarthy, M.J., Newell, S.E., Scott, J.T., Hall, N.S., Otten, T.G. and Qin, B. (2019). Mitigating eutrophication and toxic cyanobacterial blooms in large lakes: The evolution of a dual nutrient (N and P) reduction paradigm. *Hydrobiologia*, 1-17.
- Quinlan, R., A. Filazzola, A., Mahdiyan, O., Shuvo, A., Blagrove, K., Ewins, C., Moslenko, L., Gray, D., O'Reilly, C.M., Sharma, S. (In review). Relationships of total phosphorus and chlorophyll in lake worldwide. *Limnology and Oceanography*
- Reichwaldt, E. S., & Ghadouani, A. (2012). Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. *Water research*, 46(5), 1372-1393.
- Richardson, D.C., Melles, S.J., Pilla, R.M., Hetherington, A.L., Knoll, L.B., Williamson, C.E., Kraemer, B.M., Jackson, J.R., Long, E.C., Moore, K. and Rudstam, L.G. (2017). Transparency, geomorphology and mixing regime explain variability in trends in lake temperature and stratification across Northeastern North America (1975–2014). *Water*, 9(6), 442.
- Rose, K. C., Greb, S. R., Diebel, M., & Turner, M. G. (2017). Annual precipitation regulates spatial and temporal drivers of lake water clarity. *Ecological Applications*, 27(2), 632-643.
- Sanchez-Lorenzo, A., Wild, M., Brunetti, M., Guijarro, J.A., Hakuba, M.Z., Calbó, J., Mystakidis, S. and Bartok, B. (2015). Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939–2012). *Journal of Geophysical Research: Atmospheres*, 120(18), 9555-9569.
- Schindler, D. W. (1977). Evolution of phosphorus limitation in lakes. *Science*, 195(4275), 260-262.

- Schneider, P., & Hook, S. J. (2010). Space observations of inland water bodies show rapid surface warming since 1985. *Geophysical Research Letters*, 37(22).
- Sharma, S., Gray, D.K., Read, J.S., O'reilly, C.M., Schneider, P., Quadrat, A., Gries, C., Stefanoff, S., Hampton, S.E., Hook, S. and Lenters, J.D. (2015). A global database of lake surface temperatures collected by in situ and satellite methods from 1985–2009. *Scientific data*, 2(1), 1-19.
- 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.
- Sommer, U., & Lampert, W. (1997). *Limnoecology: the ecology of lakes and streams*. Oxford University Press. New York.
- Vaughan, I. P., & Gotelli, N. J. (2019). Water quality improvements offset the climatic debt for stream macroinvertebrates over twenty years. *Nature communications*, 10(1), 1-8.
- Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41(18), 6396-6402.
- Wagner, T., Soranno, P. A., Webster, K. E., & Cheruvilil, K. S. (2011). Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes. *Freshwater Biology*, 56(9), 1811-1824.
- Wetzel, R. G. (2001). *Limnology: lake and river ecosystems*. gulf professional publishing.
- Wild, M., Folini, D., Henschel, F., Fischer, N., & Müller, B. (2015). Projections of long-term changes in solar radiation based on CMIP5 climate models and their influence on energy yields of photovoltaic systems. *Solar Energy*, 116, 12-24.
- Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., & Sharma, S. (2020). Global lake responses to climate change. *Nature Reviews Earth & Environment*, 1-16.
- Wurtsbaugh, W. A., Paerl, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water*, 6(5), e1373.

Chapter 1: Total phosphorus and climate are equally important predictors of water quality in lakes: A global analysis

Arnab Shuvo*¹, Catherine M. O'Reilly*², Kevin Blagrave¹, Carolyn Ewins¹, Alessandro Filazzola¹, Derek Gray³, Octavia Mahdiyan¹, Luke Moslenko¹, Roberto Quinlan¹, and Sapna Sharma¹

¹Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada

²Department of Geography, Geology, and the Environment, Illinois State University, Normal, Illinois, USA

³Department of Biology, Wilfrid Laurier University, Waterloo, Ontario, Canada

***Joint first authors**

Keywords: Chlorophyll, Phosphorus, Climate Change, Morphology, Lakes, Water Quality

Chapter 1 was submitted to Aquatic Sciences.

Shuvo, A., O'Reilly, CM., Blagrave, K., Ewins, C., Filazzola, A., Gray, D., Mahdiyan, O., Moslenko, L., Quinlan, R., Sharma, S. (In review). Total phosphorus and climate are equally important predictors of water quality in lakes: A global analysis. Aquatic Sciences.

MY CONTRIBUTION

The manuscript titled, “Total phosphorus and climate are equally important predictors of water quality in lakes: A global analysis” was submitted to the journal *Aquatic Sciences*. The authors of this publication include Catherine M. O’Reilly, Kevin Blagrove, Carolyn Ewins, Alessandro Filazzola, Derek Gray, Octavia Mahdiyan, Luke Moslenko, Roberto Quinlan, and Sapna Sharma. I analyzed all of the data, and wrote the Abstract, Introduction, Methods, Results, and Discussion. I also created figures, and supplementary figures and tables.

SUMMARY

Water quality degradation is one of the largest threats to freshwater ecosystems. Nutrient inputs, land use changes, and climate are expected to be the most important drivers of water quality degradation. Here, we quantify the relative importance of nutrient inputs, climate, and lake geomorphometry as drivers of primary production in freshwater lakes globally, using chlorophyll as a proxy. We synthesized a global database on lake water chemistry, with a focus on chlorophyll, in addition to lake geomorphometry and climate for up to 3878 freshwater lakes around the globe. We found that globally, total phosphorus (TP) was the most influential driver, explaining 40% of the variation in chlorophyll concentrations. A combination of climate variables, including spring solar radiation, summer temperature, and spring precipitation, explained an additional 40% of the variation in chlorophyll. Lake geomorphology only explained 20% of the variation in chlorophyll with mean depth being the most influential geomorphometric driver. Although there are increased efforts and regulations in place for land use and farming, nutrient inputs continue to be the leading cause of primary production in lakes. However, the influence of climatic variables acting synergistically (temperature, precipitation, cloud cover and solar radiation) are nearly equal to that of total phosphorus, suggesting nutrient management efforts are not sufficient alone to mitigate water quality degradation. Our findings underscore the critical need to incorporate climate factors into water quality management given current climate change.

INTRODUCTION

Eutrophication is one of the most prevalent threats to freshwater resources, with consequences for public health, food security, biodiversity, and other ecosystem services, highlighting the urgency of understanding controls on water quality. Nutrients, particularly total phosphorus, have long been recognized as a primary factor influencing water quality (Quinlan et al., submitted; Schindler, 1977). Hydrology and climate are tightly coupled and climate change is affecting water quality through warming temperatures, changing precipitation patterns, and increases in nutrient inputs from land use (Michalak, 2016). For example, regions that are experiencing increases in precipitation are likely to see a rise in eutrophication driven by increases in nutrient runoff (Sinha et al., 2017). These increases in algal biomass have further implications for greenhouse gas sources (Beaulieu et al., 2019) and sinks (Webb et al., 2019), as well as for the presence of cyanobacterial toxins (Hayes and Vanni, 2018). Understanding how climate drivers interact with nutrients to impact water quality is important for environmental management and improving water quality can subsequently mitigate the consequences of climate change on aquatic ecosystems (Vaughan et al., 2019).

Climate is increasingly being recognized as an important driver of water quality at local to regional scales. Synchronous changes in lake algal biomass associated with climate-mediated variability in the hydrological cycle suggest that climate plays an important role in chlorophyll dynamics (Baines et al., 2000). In general, increases in nutrient loading from higher precipitation exacerbate water quality problems through increases in chlorophyll, changes in phytoplankton community composition (Jeppesen et al., 2009), and decreases in water clarity (Rose et al., 2017). In the northeastern United States, summer air temperatures and winter precipitation were

significant predictors of summer algal biomass (Collins et al., 2019). Temperature has been associated with increased chlorophyll in other regional studies (Woelmer et al., 2016). The dominance of cyanobacteria, which may generate harmful algal blooms (HABs), increases at warmer temperatures despite no changes to nutrient levels or total nitrogen:total phosphorus ratios (McQueen and Lean, 1987). Warmer air temperatures contribute to earlier ice-out, leading to a longer growing season and higher chlorophyll *a* levels (Preston et al., 2016), and changes in the onset and strength of thermal stratification can lead to increases (e.g. Jöhnk et al., 2008) or decreases in algal biomass (e.g. Verburg and Hecky, 2003). Additionally, while temperature is generally positively correlated with algal biomass, the response can vary depending on lake trophic status, trophic interactions, and resource availability (Kraemer et al., 2017).

Solar radiation is important for primary production. Broadly, the earth's surface underwent a period of declining solar radiation from around 1950 until the mid-1980s, referred to as 'global dimming', after which solar radiation increased and stabilized (Wild et al., 2005; Sanchez-Lorenzo et al., 2015), creating a recent period of 'global brightening' (Wild et al., 2015). These large-scale changes have been attributed to changes in the transparency of the atmosphere owing to anthropogenic aerosol emissions and/or cloudiness (Gan et al., 2014; Long et al., 2009; Mateos et al., 2014). Recent trends in brightening are most apparent in regions that have long been industrialized, whereas newly industrialized regions have seen continued dimming (Wild, 2012). For example, from 1980 to 2010 in Europe and North America, average surface solar radiation increased up to 8 W m^{-2} per decade (Wild, 2012). In contrast, surface solar radiation decreased by up to 10 W m^{-2} per decade in India (Wild, 2012). Patterns in solar radiation brightening and dimming vary spatially (e.g. Wild et al., 2012) and temporally (e.g. Sanchez-

Lorenzo et al., 2015). For example there may be increases in solar radiation in one season but decreases in another (Deng et al., 2019; Sanchez-Lorenzo et al., 2015). Long term trends can be substantially greater than that of the 11-yr solar sunspot cycle, which varies only by approximately 1 W m^{-2} and is known to have a wide range of effects (Engels and van Geel, 2012). Thus, these changes in solar radiation have the potential to influence algal growth and biomass.

Lake basin morphology and hydrology have the capacity to influence water quality, particularly with respect to interactions with other drivers. Factors such as lake depth and surface area influence physical, chemical and biological processes within the lake, with consequences for mixing regimes, nutrient availability, and residence times. Deeper lakes typically have both lower total phosphorus (TP) and lower chlorophyll concentrations (Canfield et al., 2016; Fergus et al., 2016; Liu et al. 2012). For example, deeper lakes had clearer water in Wisconsin, although this only occurred in watersheds with low levels of agriculture (Rose et al., 2017). Lake surface area has also been identified as an important contributor to chlorophyll (Woelmer et al., 2016). Among deeper lakes in eastern China, those with a smaller surface area had higher chlorophyll concentrations compared to larger lakes (Huang et al., 2014). Longer residence times have been associated with lower chlorophyll concentrations in Europe (Noges, 2009), but with higher concentrations in eastern China (Liu et al., 2012). However, none of these previous studies have comprehensively considered lake geomorphometric and hydrologic characteristics in the context of both climate and nutrient factors.

Spatial patterns in water quality can be highly variable, driven by many interacting factors (Fergus et al., 2016). Current trajectories in climate, the continuing influence of nutrient management legacies (Sharpley et al., 2013), and the growing demand for clean freshwater underscore the importance of understanding drivers of water quality. Nutrient concentrations, particularly phosphorus, are well established as key predictors of chlorophyll (Abell et al., 2012; McCauley et al., 1989; Quinlan et al., in review), but the relative role of other factors has focused primarily on temperature and precipitation (e.g. Collins et al., 2019) and at the regional scale (e.g. Collins et al., 2019). This excludes the potential impact of climate factors that have large spatial patterns, such as solar radiation (Wild, 2012) and other variables that can influence solar radiation at more local scales (e.g. cloudiness). We assembled a global database of chlorophyll concentrations from 3878 lakes in 41 countries to explore predictors and spatial patterns of water quality. In this study we asked, in addition to total phosphorus, which climatic and lake geomorphometric factors influence water quality? We examined the relative importance of nutrient inputs, climate conditions, and lake morphology factors on chlorophyll *a* concentrations for lakes worldwide. Within climate factors, we included the relative roles of temperature, solar radiation, cloud cover, and precipitation, at various seasonal scales.

METHODS

2.1 Data Acquisition and Dataset Compilation

We conducted a systematic review to develop a global chlorophyll database for approximately 10,000 lakes with over 25,000 sampling points (Filazzola et al., in revision). For this study, we used an overall global dataset which included both chlorophyll and total phosphorus data for 3,878 lakes distributed across 41 countries (Figure 1; Table S1). Each chlorophyll *a* measurement corresponded with the year in which the measurement was taken and the lake's latitude and longitude. In instances where the same lake was sampled in multiple locations, we took the median values of measured data for that lake. When a lake was sampled multiple times within the same year (e.g. monthly), we extracted data from the time when chlorophyll *a* values were greatest. If a lake was sampled over multiple years, we selected data from the most recent year because it most closely resembles the most recent status of that lake. TP and climate data were collected in the same manner and matched with corresponding chlorophyll values. The Laurentian Great Lakes were also excluded because of significant spatial heterogeneity in water quality due to their size. While this dataset includes samples collected with a variety of methodologies, we do not anticipate that the international multi-jurisdictional nature of the data will influence analyses, as Hanna and Peters (1991) showed that differences in protocol and sampling depth for TP and chlorophyll *a* did not add variation to TP-chlorophyll *a* response models (Filazzola et al., in review; Quinlan et al., in review).

Although most climate data were available for all lakes, solar radiation data were only available for 2561 lakes in 7 countries (Figure 1, Table S1). Shortwave incoming solar radiation in the near infrared (0.2-4 μm) data were collected from CLARA A-1 which is gridded at a 0.25 x 0.25

degrees resolution between 1982 and 2009 (Karlsson et al., 2013). For the other climate variables, we extracted air temperature (°C), precipitation (mm), cloud cover (%) for each lake and sampling year from the University of East Anglia's Climate Research Unit (CRU) online open access database (Harris et al., 2020). CRU interpolates air temperature and precipitation from meteorological station measurements and summarizes the database on a grid at a 0.5° latitude and longitude resolution. Using the monthly mean temperature, precipitation, cloud cover and solar radiation values, we calculated the mean, maximum, and standard deviation annually, for each season, and at a lag of 1 year. For the scope of our study we used spring (March, April, May for the Northern Hemisphere, and September, October, November for the Southern Hemisphere) and summer (June, July, August for the Northern Hemisphere, and December, January, February for the Southern Hemisphere) climate variables because that is when algal biomass is the most abundant. Geomorphic data for each lake such as: mean depth, lake area, residence time, volume, elevation, and watershed area were taken from HydroLAKES (Messenger et al. 2016)

2.2 Data analyses

Our dataset with solar radiation (D+SR) comprised of 2561 lakes where chlorophyll (response variable) was tested against total phosphorus (TP), climate (solar radiation, summer and spring air temperature, precipitation, and cloud cover) and geomorphometry (lake mean depth, elevation, area, residence time, total volume, and watershed area). The primary variables of interest (Chl-a and TP) were log₁₀ transformed before applying the randomforest function (package = randomforest) with iteration = 5 (Liaw and Wiener, 2002). Because we had a larger dataset of 3878 lakes that included all variables except solar radiation, we also conducted all of the same analyses on this larger dataset (Dataset excluding solar radiation; D-SR).

2.3 Drivers of Water Quality

We assessed normality of all variables using a Kolmogorov-Smirnov test and a qqplot using the “*ks.test* and *qqnorm*” functions in R. Chlorophyll *a* and total phosphorus (TP) were log-transformed to satisfy assumptions of normality. Kendall’s Tau correlation coefficients were calculated between all variables using the “*cor*” function in R. We used i) random forests; ii) linear models; and iii) principal component analysis to evaluate the relationships between chlorophyll and nutrient, climatic, and geomorphometric drivers,

We conducted a random forest analysis to identify the nutrient, climatic, and geomorphometric drivers of chlorophyll *a*. Decision trees are the basic building blocks for a random forest. The decision tree is a machine learning algorithm which fits complex datasets with interacting predictor variables (Ali et al., 2012; Ben-Haim and Tom-Tov, 2010). Random forest analyses are classified as a strong learner because they are built as an ensemble of decision trees (Ali et al., 2012; Breiman, 2001). Random forest analyses work through bagging or bootstrap aggregation where subsets of the training data are randomly sampled, and a model is fit to these smaller data sets after which the predictions are aggregated (Ali et al., 2012; Breiman, 1996, 2001). For our analysis we conducted 5 iterations consisting of 500 trees and proximity = TRUE as a similarity measure which accounts for the number of times certain variables end up on the same node. Cross validation is not required because bootstrap sampling excludes data that was not used in the model and creates an out of bag error (oob). This prevents overfitting in the random forest model.

We then investigated the relationship between Chl-a and TP, and the climate variables mentioned previously. We fit Chl-a (response variable) against TP and climate variables (explanatory variables) in linear models. The Linear regression models the relationship between the predictor and response variables by fitting a linear equation ($Y = X\beta + \epsilon$) to observed data (Bhattacharya and Burman, 2016).

To test the drivers of chlorophyll, we fitted linear regression models (function: `lm`) with chlorophyll as the response variable and the climate, nutrient, and lake morphometric variables as the predictors. We ran separate models for each of the following predictors. Finally, we also calculated the Akaike information criterion (AIC) values for each corresponding model - where a lower AIC score is indicative of the most parsimonious model.

We conducted a Principal Components Analysis (PCA) in R (function `prcomp`) to summarize the relationships between chlorophyll, climate and nutrient levels with a reduced number of dimensions. The PCA is an algorithm that generates principal components (axes) along which the data show the maximum amount of variation. One of the outputs of a PCA is a two-dimensional ordination plot which can allow for the visualization of relationships among predictor variables and among individual observations (Legendre and Legendre, 2012). We did not include the geomorphometric variables because we did not want to overpopulate and complicate the model. We also generated the PCA visual using the '`fviz_pca_var()`' function from the package = `factoextra`.

RESULTS

3.1 Drivers of Water Quality including Solar Radiation (using D+SR)

A random forest analysis on D+SR using nutrient concentrations, climatic variables, and lake geomorphometric characteristics as predictors explained 60% of variation in lake chlorophyll levels globally. The random forest regression had a mean of squared residual of 0.12 suggesting that the model is a good fit (Figure 2a). Total phosphorus was the most important driver of lake chlorophyll concentrations in D+SR explaining approximately 42% of the total variation (Figure 2a). Other climatic variables, comprising summer and spring air temperatures, solar radiation, precipitation, and cloud cover cumulatively explained 38% of the variation. Notably, the most influential climatic variables were summer and spring air temperature, but spring solar radiation was the next most influential climate variable - even more so than precipitation and cloud cover. Geomorphometric characteristics accounted for the least amount of variation in chlorophyll explaining 20% of the variation in D+SR, with lake watershed area being the most important (Figure 2a).

The linear models on D+SR supported the results from the random forest (Figure 3). The TP-Chl-a relationship was best explained by a linear model ($P < 0.001$, $R^2_{\text{adj}} = 0.47$, $\text{AIC} = 6581.4$) which suggests a fairly strong model fit and a significant positive relationship (Figure 3b). The linear models also supported the influential role of solar radiation ($P < 0.001$, $R^2_{\text{adj}} = 0.041$, $\text{AIC} = 8367.04$) (Figure 3a) and summer air temperature ($P < 0.001$, $R^2_{\text{adj}} = 0.14$, $\text{AIC} = 8085.00$), and summer precipitation ($P < 0.001$, $R^2_{\text{adj}} = 0.019$, $\text{AIC} = 8425.50$) as positive drivers of Chl-a.

Conversely, summer cloud cover ($P < 0.001$, $R^2_{\text{adj}} = 0.045$, $\text{AIC} = 8356.17$) was negatively associated with chlorophyll a .

The first two axes of the PCA of D+SR explained 53.7% of the variation, with principal axis one explaining 33.8% and principal axis two explaining 19.9% of the variation (Figure 5a). Spring solar radiation is closely associated with chlorophyll suggesting that spring solar radiation has a direct and significant influence on chlorophyll growth. Spring solar radiation is also associated with air temperature and TP, suggesting its potential influence on these variables (Figure 5a, Figure S1)

3.2 Drivers of Water Quality excluding Solar Radiation

Because surface solar radiation data availability was limited to the time period before 2009, and we had more recent chlorophyll data from an additional 1,317 lakes, we repeated these analyses on the larger dataset excluding solar radiation (D-SR; $n=3,878$).

The random forest model using D-SR produced similar results to the analysis using D+SR but explained a lower amount of variation (Figure 2b). The model explained 55.2% of variation with a mean of squared residual of 0.16. Once again, TP was the dominant driver of chlorophyll explaining 38% of the variation, while climate variables (with solar radiation excluded) explained 37% of the variation. Of the climatic variables using D-SR, temperature explained the most variation, followed by cloud cover, and precipitation. Geomorphometric variables were again the least influential, explaining 25% of the variation, and lake mean depth being most important.

The linear models for TP, climate, and geomorphometric variables remained similar when using D-SR. TP was the most influential driver of Chl-a, but seasonal (spring and summer) climate variables (air temperature, precipitation and cloud cover) were also highly influential, and mean depth emerged as the most influential geomorphometric variable (Figure 4). The relationship between TP and Chl using D-SR ($P < 0.001$, $R^2_{adj} = 0.43$, AIC = 11243.92) was once again significant and positive (Figure 4a). Of the climate variables, summer air temperature ($P < 0.001$, $R^2_{adj} = 0.13$, AIC = 12850.85) and precipitation ($P < 0.001$, $R^2_{adj} = 0.0094$, AIC = 13355.72) were positive drivers of Chl-a. While summer cloud cover ($P < 0.001$, $R^2_{adj} = 0.023$, AIC = 13303.09) and mean depth ($P < 0.001$, $R^2_{adj} = 0.062$, AIC = 3920.47) were negative drivers of chlorophyll.

The PCA calculated using D-SR explained 53.2% of the variation, where the first and second axis explained 30.6% and 22.6% of the variation, exclusively. In the PCA spring solar radiation and cloud cover were the dominant drivers of the first axis (Figure 5b). The second axis was explained by precipitation in the spring and summer (Figure 5b). The PCA using D-SR exhibited similar patterns to the PCA using D+SR, where TP and air temperature were positively correlated to chlorophyll concentrations and negatively related to summer and spring cloud cover (Figure 5b; Figure S2). Spring and summer climatic conditions were strongly positively correlated suggesting that both seasons strongly influenced summer chlorophyll conditions.

DISCUSSION

Our study examined chlorophyll *a* concentrations globally to identify the most influential drivers of primary production and attempted to understand how these drivers may synergistically influence the abundance of lake algae. Our models confirmed that total phosphorus (TP) was the most influential driver of chlorophyll concentrations. However, the cumulative impact of climate variables (temperature, solar radiation, precipitation, and cloud cover) was equally important as TP for chlorophyll concentrations in lakes. In particular, solar radiation had a critical influence on summer chlorophyll when incorporated into our models. Although models were poorer when solar radiation was excluded, climate variables still explained a significant proportion of variation in chlorophyll levels. Lake geomorphometric characteristics also explained some variation in our models, but to a much lesser extent when compared to TP and climate variables.

Our results underscore the role of total phosphorus as a primary predictor of chlorophyll in freshwater lakes. We found that TP explained approximately 40% of the variation in chlorophyll *a* concentrations globally. Total phosphorus has been well established as a predictor of chlorophyll, including at global scales (Abell et al., 2012; McCauley et al., 1989; Quinlan et al., in review). Freshwater ecosystems naturally act as nutrient sinks, as well as receive phosphorus inputs from a number of human activities (e.g. excess P-fertilizer use (Carpenter, 2005)), P-containing pesticides (Arbuckle and Downing, 2001), and domestic and industrial sewage (Smith and Schindler, 2009). Our results showing that the correlation between TP and chlorophyll concentration was moderate (0.5; Supporting Information) and that TP explained approximately 40% of the variation in chlorophyll concentrations are consistent with previous examinations of the TP-chlorophyll relationship at global scales (Abell et al. 2012; Quinlan et al., in review).

Across a large gradient of phosphorus concentrations, the TP-chlorophyll relationship is sigmoidal, structured by breakpoints related to limitation by nitrogen among other factors (Quinlan et al., in review). Variation within the global TP-chlorophyll relationship is large and the nature of the relationship varies among regions (Filstrup et al., 2014). This variation is postulated to be due to a wide array of factors in addition to nitrogen limitation both when phosphorus is very high (e.g. Filstrup and Downing, 2017) and very low (e.g. Morris and Lewis, 1988), as well as watershed land use (Filstrup et al., 2014). For example, the accumulation of algal biomass could be limited by cold water temperatures (Butterwick et al., 2005) or shading from dense algal blooms (Agustí et al., 1990). Alternatively algal biomass could be greater than expected due to a combination of lake depth and transparency allowing for specific species to thrive (Hennemann and Petrucio, 2016).

In addition to exploring the chlorophyll-TP relationship, we were able to examine the importance of climate variables for predicting chlorophyll levels. The combination of climate variables considered in our study explained an additional ~40% of the variation in global chlorophyll *a* concentrations when calculated using D+SR and D-SR. For example, we found that chlorophyll *a* concentrations were higher in regions with greater solar radiation inputs. Light limitation of algal growth has been demonstrated in many lakes by experiments (e.g. Dubourg et al., 2015) and can vary by season (e.g. Kolzau et al., 2014). For example, in western Lake Erie, solar radiation was a top predictor of chlorophyll concentrations in spring and summer compared to other climate and nutrient factors, explaining 38 and 25% of the total variation respectively each season (Tian et al., 2017). Increases in winter and spring solar radiation over the past two decades were drivers of phytoplankton abundance and composition in those seasons in Lake

Taihu (Deng et al., 2019). Increases in solar radiation have also been linked to warming of lake water temperatures around the globe (O'Reilly et al., 2015; Schmid and Köster, 2016; Zhong et al., 2016). We observed strong correlations between solar radiation and air temperature across a global set of lakes (Figure S1). Solar radiation contributes to lake heat budgets and water temperatures (Schmid et al., 2014), thus our lake temperatures were also likely warmer with greater solar radiation. Warmer water temperatures increase algal growth rates and have been linked to increases in chlorophyll and algal blooms (Ho et al., 2019). Since we did not have water temperature data, we could not examine if the importance of solar radiation for predicting chlorophyll was related to water temperature, increased light availability, or both. Additionally, we were unable to explore trends in chlorophyll concentrations within lakes, which could also facilitate understanding the relative roles of these different drivers. We suggest that future studies investigate the contribution of global brightening as well as water temperature to chlorophyll trends in lakes around the world.

Considering climate factors other than solar radiation (using D-SR), our results at the global scale were similar to those found in regional studies that incorporated climate at seasonal scales. Our analyses suggested that the largest driver of summer chlorophyll was summer temperature, followed by spring solar radiation and spring precipitation (Figure 2a). Chlorophyll concentrations were highest in warmer regions with higher solar radiation inputs (Figure 3a; Figure S1). When analyzing D-SR, cloud cover became one of the top three factors, which was highly negatively correlated to solar radiation in this study. Warmer water temperatures facilitate faster growth rates, and spring precipitation likely increases nutrient inputs to lakes, potentially setting up ideal conditions for algal growth into the summer. However, the selection of climate

variables found to predict chlorophyll levels and their relative importance can vary significantly among studies. For example, summer and spring temperature, as well as winter precipitation, were the most important predictors of chlorophyll among lakes in the northeastern region of the United States (Collins et al., 2019). However, a similar study found that water clarity in the northeast US correlated more strongly with summer precipitation, and not with maximum air temperatures, while in the midwest there was no relationship to summer precipitation and a negative correlation with maximum air temperatures (McCullough et al., 2019). The ability to identify the major drivers of lake chlorophyll depends on both the suite of factors incorporated into the analyses as well as the time scales that are involved. Most prior studies have been regional in scale and investigated only temperature and precipitation (e.g. Collins et al., 2019) using lag times of conditions ranging between the prior two weeks (Lennard, 2019) to the previous year (Collins et al., 2019). In addition, lake trophic status can be an important consideration, as a global study of 188 lakes found that the relationship between water temperature and chlorophyll was positive for eutrophic lakes, but negative for oligotrophic ones (Kraemer et al., 2017). Thus, while it is difficult to compare specific findings across studies, temperature and precipitation are consistently key overarching drivers, and it is clear that lake water quality is sensitive to climate in ways that are highly variable within and across regions.

Our results showed that the relationships between chlorophyll and lake geomorphometric and hydrological factors were less important in the broader context of climate and nutrients.

Although geomorphic factors (area, mean depth, and elevation) and hydrology (residence time) did play a role in determining chlorophyll, these factors were much less important than climate and total phosphorus. Lake mean depth was the most important of these factors, but in

combination, morphology and residence time only contributed to about 20% of the variation in chlorophyll (Figure 1). Mean depth and water residence time have explained a much greater proportion of the variation in other studies, but this may be because all lakes considered in those studies were relatively deep (Liu et al., 2012). When lakes are deep, other factors such as surface area can also become important (Huang et al., 2014). Ultimately, lake depth affects the mixing regime and degree of stratification, which is probably the more proximal factor influencing the TP-chlorophyll relationship. Chlorophyll was negatively correlated with mean depth, as has been found in other studies (Figure 3f, Canfield et al., 2016; Fergus et al., 2016). Longer residence times allow for the development of greater biomass in lakes that already have high nutrient concentrations (e.g. Huang et al., 2014), but do not appear to influence chlorophyll concentrations in lakes that have low nutrient concentrations (e.g. Nöges, 2009). Overall, our results are consistent with various regional studies that indicate the influence of lake morphology and hydrology on chlorophyll but highlight that these factors are relatively minor compared to climate and total phosphorus.

CONCLUSION

While our study underscored the importance of TP as a driver of chlorophyll concentrations, we also found that the role of climate was just as important as TP. The primary role of TP in chlorophyll has been well established, but the variation within this TP-chlorophyll relationship is large (Quinlan et al., submitted). Our results indicate that the factors contributing to this variation are primarily composed of a combination of climate and geomorphometric drivers, but climate variables clearly emerged as important drivers of chlorophyll, highlighting their role in ecosystem structure at various spatiotemporal scales (Collins et al., 2019). Geomorphometric and hydrological features were also important but had a reduced role relative to climate in terms of explaining overall variation in summer chlorophyll. The relatively similar contributions of climate and nutrients suggest that management plans focused primarily on nutrient reductions may also need to consider the implications of ongoing climate change in order to achieve the desired outcomes (Trolle et al., 2011; Paerl et al., 2016). Our study highlights the specific aspects of climate that should be taken into consideration, and fortunately future projections for these climate variables are readily available from existing climate models (Flato et al., 2013). Changes in temperature and precipitation during spring and summer seasons are occurring across many parts of the world, and these changes will clearly present challenges for management of water quality if not incorporated into water quality management frameworks.

ACKNOWLEDGMENTS

Data are available in Filazzola et al. (in revision). We thank the numerous data providers who provided data for this study. Funding for this study was provided by the Ontario Ministry of Innovation Early Researcher Award, York Research Chair and Natural Sciences Engineering Research Council Discovery Grant to SS, and the Ontario Ministry of Environment and Climate Change Best in Science Grant to DG and SS. SS and DG secured funding for the project; SS led the project. AF, KB, AS, LM, CE, and OM assembled the dataset; AS, AF, and DG conducted analyses; KB and AS made figures; SS, COR, AS, OM, and DG wrote sections of the text. All authors contributed ideas, editing and revision. Authors between the co-leads and last author have similar contributions and are listed in alphabetical order.

REFERENCES

- Abell, J. M., Özkundakci, D., Hamilton, D. P., & Jones, J. R. (2012). Latitudinal variation in nutrient stoichiometry and chlorophyll-nutrient relationships in lakes: A global study. *Fundamental and Applied Limnology*, 181(1), 1–14. doi:10.1127/1863-9135/2012/0272
- Agusti, S., Duarte, C. M., & Canfield, D. E. (1990). Phytoplankton abundance in Florida lakes: Evidence for the frequent lack of nutrient limitation. *Limnology and Oceanography*, 35(1), 181–187. doi:10.4319/lo.1990.35.1.0181
- Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues*, 9(5), 272–278.
- Arbuckle, K. E., & Downing, J. A. (2001). The influence of watershed land use on lake N : P in a predominantly agricultural landscape. *Limnology and Oceanography*, 46(4), 970–975. doi:10.4319/lo.2001.46.4.0970
- Baines, S. B., Webster, K. E., Kratz, T. K., Carpenter, S. R., & Magnuson, J. J. (2000). Synchronous behavior of temperature, calcium, and chlorophyll in lakes of northern Wisconsin. *Ecology*, 81(3), 815–825. doi:10.1890/0012-9658(2000)081[0815:SBOTCA]2.0.CO;2
- Beaulieu, J. J., DelSontro, T., & Downing, J. A. (2019). Eutrophication will increase methane emissions from lakes and impoundments during the 21st century. *Nature Communications*, 10(1), 3–7. doi:10.1038/s41467-019-09100-5
- Ben-Haim, Y., & Tom-Tov, E. (2010). A streaming parallel decision tree algorithm. *Journal of Machine Learning Research*, 11, 849–872
- Bhattacharya, P. K., & Burman, P. (2016). Theory and methods of statistics. Academic Press.
- Breiman, L. (1996). Bagging Predictors, URL: <https://link.springer.com/article/10.1007%2FBF00058655>. *Machine Learning*, 24(421), 123–140. doi:10.1007/BF00058655
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. doi:10.1201/9780429469275-8
- Butterwick, C., Heaney, S. I., & Talling, J. F. (2005). Diversity in the influence of temperature on the growth rates of freshwater algae, and its ecological relevance. *Freshwater Biology*, 50(2), 291–300. doi:10.1111/j.1365-2427.2004.01317.x
- Canfield, D. E., Bachmann, R. W., Stephens, D. B., Hoyer, M. V., Bacon, L., Williams, S., et al. (2016). Monitoring by citizen scientists demonstrates water clarity of Maine (USA) lakes is stable, not declining, due to cultural eutrophication. *Inland Waters*, 6(1), 11–27. doi:10.5268/IW-6.1.864
- Carpenter, S. R. (2005). Eutrophication of aquatic ecosystems: Bistability and soil phosphorus. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10002–10005. doi:10.1073/pnas.0503959102
- Collins, S. M., Yuan, S., Tan, P. N., Oliver, S. K., Lapierre, J. F., Cheruvilil, K. S., et al. (2019). Winter Precipitation and Summer Temperature Predict Lake Water Quality at Macroscales. *Water Resources Research*, 55(4), 2708–2721. doi:10.1029/2018WR023088
- Deng, C., Tong, Y., Chen, L., Yuan, W., Sun, Y., Li, J., et al. (2019). Impact of particle chemical composition and water content on the photolytic reduction of particle-bound mercury. *Atmospheric Environment*, 200, 24–33. doi:10.1016/j.atmosenv.2018.11.054
- Dubourg, P., North, R. L., Hunter, K., Vandergucht, D. M., Abirhire, O., Silsbe, G. M., et al. (2015). Light and nutrient co-limitation of phytoplankton communities in a large reservoir:

- Lake Diefenbaker, Saskatchewan, Canada. *Journal of Great Lakes Research*, 41, 129–143. doi:10.1016/j.jglr.2015.10.001
- Engels, S., & van Geel, B. (2012). The effects of changing solar activity on climate: contributions from palaeoclimatological studies. *Journal of Space Weather and Space Climate*, 2, A09. doi:10.1051/swsc/2012009
- Fergus, C. E., Finley, A. O., Soranno, P. A., & Wagner, T. (2016). Spatial variation in nutrient and water color effects on lake chlorophyll at macroscales. *PLoS ONE*, 11(10), 1–20. doi:10.1371/journal.pone.0164592
- Filazzola, A., Mahdiyna O., Shuvo, A., Ewins, C., Moslenko, L., Sadid, T., et al. (in revision). A global database of chlorophyll and water chemistry in freshwater lakes. *Scientific Data*.
- Filstrup, C. T., & Downing, J. A. (2017). Relationship of chlorophyll to phosphorus and nitrogen in nutrient-rich lakes. *Inland Waters*, 7(4), 385–400. doi:10.1080/20442041.2017.1375176
- Filstrup, C. T., Wagner, T., Soranno, P. A., Stanley, E. H., Stow, C. A., Webster, K. E., & Downing, J. A. (2014). Regional variability among nonlinear chlorophyll-phosphorus relationships in lakes. *Limnology and Oceanography*, 59(5), 1691–1703. doi:10.4319/lo.2014.59.5.1691
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., et al. (2013). Evaluation of Climate Models. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Gan, C. -M., Pleim, J., Mathur, R., Hogrefe, C., Long, C. N., Xing, J., et al. (2014). Assessment of the effect of air pollution controls on trends in shortwave radiation over the United States from 1995 through 2010 from multiple observation networks. *Atmospheric Chemistry and Physics*, 14, 1701–1715.
- Hanna, M., & Peters, R. H. (1991). Effect of sampling protocol on estimates of phosphorus and chlorophyll concentrations in lakes of low to moderate trophic status. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(10), 1979–1986.
- Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7, 109 (2020). doi.org/10.1038/s41597-020-0453-3
- Hayes, N. M., & Vanni, M. J. (2018). Microcystin concentrations can be predicted with phytoplankton biomass and watershed morphology. *Inland Waters*, 8(3), 273–283. doi:10.1080/20442041.2018.1446408
- Hennemann, M. C., & Petrucio, M. M. (2016). High chlorophyll a concentration in a low nutrient context: discussions in a subtropical lake dominated by cyanobacteria. *Journal of Limnology*, 75(3). doi.org/10.4081/jlimnol.2016.1347
- Ho, J.C., Michalak, A.M., & Pahlevan, N., (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, 574, 667–670. doi:10.1038/s41586-019-1648-7.
- Huang, J., Xu, Q., Xi, B., Wang, X., Jia, K., Huo, S., et al. (2014). Effects of lake-basin morphological and hydrological characteristics on the eutrophication of shallow lakes in eastern China. *Journal of Great Lakes Research*, 40(3), 666–674. doi.org/10.1016/j.jglr.2014.04.016.

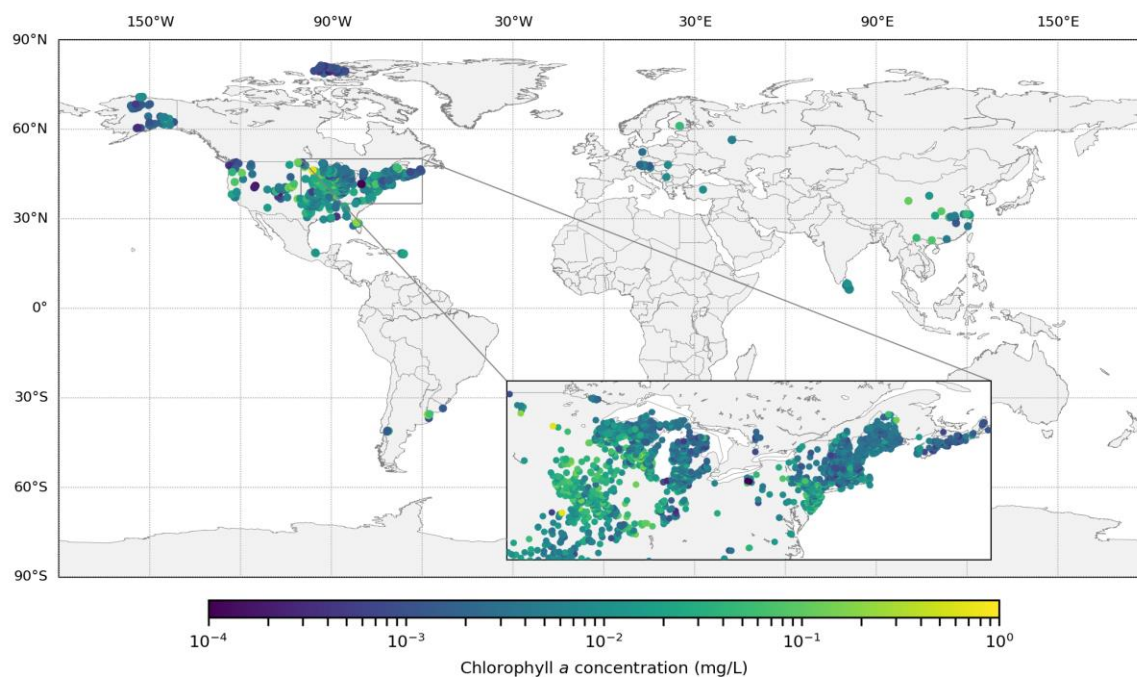
- Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K. M., Andersen, H. E., et al. (2009). Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. *Journal of Environmental Quality*, 38(5), 1930–1941. doi:10.2134/jeq2008.0113.
- Jöhnk, K. D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P. M., & Stroom, J. M. (2008). Summer heatwaves promote blooms of harmful cyanobacteria. *Global Change Biology*, 14(3), 495–512. doi:10.1111/j.1365-2486.2007.01510.x.
- Karlsson, K. G., Riihelä, A., Müller, R., Meirink, J. F., Sedlar, J., Stengel, M., et al. (2013). CLARA-A1: A cloud, albedo, and radiation dataset from 28 yr of global AVHRR data. *Atmospheric Chemistry and Physics*, 13(10), 5351–5367. doi:10.5194/acp-13-5351-2013.
- Kolzau, S., Wiedner, C., Rücker, J., Köhler, J., Köhler, A., & Dolman, A. M. (2014). Seasonal patterns of Nitrogen and Phosphorus limitation in four German Lakes and the predictability of limitation status from ambient nutrient concentrations. *PLoS ONE*, 9(4). doi:10.1371/journal.pone.0096065.
- Kraemer, B. M., Mehner, T., & Adrian, R. (2017). Reconciling the opposing effects of warming on phytoplankton biomass in 188 large lakes. *Scientific Reports*, 7(1), 1–7. doi:10.1038/s41598-017-11167-3
- Legendre, P., & Legendre, L. F. (2012). *Numerical ecology*. Elsevier.
- Lennard, C. (2019). Multi-Scale Drivers of the South African Weather and Climate. In *The Geography of South Africa* (pp. 81-89). Springer, Cham.
- Liaw, A., & M. Wiener. (2002). *Classification and Regression by randomForest*. *R News* 2/3: 18–22.
- Liu, W., Li, S., Bu, H., Zhang, Q., & Liu, G. (2012). Eutrophication in the Yunnan Plateau lakes: the influence of lake morphology, watershed land use, and socioeconomic factors. *Environmental Science and Pollution Research*, 19(3), 858-870.
- Long, C. N., Dutton, E. G., Augustine, J. A., Wiscombe, W., Wild, M., McFarlane, S. A., & Flynn, C. J. (2009). Significant decadal brightening of downwelling shortwave in the continental United States, *Journal of Geophysical Research*, 114, D00D06. doi:10.1029/2008JD011263.
- Mateos, D., Sanchez-Lorenzo, A., Antón, M., Cachorro, V. E., Calbó, J., Costa, M. J., et al. (2014). Quantifying the respective roles of aerosols and clouds in the strong brightening since the early 2000s over the Iberian Peninsula, *Journal of Geophysical Research: Atmos.*, 119, 10,382–10,393. doi:10.1002/2014JD022076.
- McCauley, E., Downing, J. A., & Watson, S. (1989). Sigmoid relationships between nutrients and chlorophyll among lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 46(7), 1171–1175. doi:10.1139/f89-152
- McCullough, I. M., Cheruvilil, K. S., Collins, S. M., & Soranno, P. A. (2019). Geographic patterns of the climate sensitivity of lakes. *Ecological Applications*, 29(2), 1–14. doi:10.1002/eap.1836
- McQueen, D. J., & Lean, D. R. S. (1987). Influence of water temperature and nitrogen to phosphorus ratios on the dominance of blue-green algae in Lake St. George, Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 44(3), 598–604. doi:10.1139/f87-073
- Messenger, 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*: 13603. doi:10.1038/ncomms13603

- Michalak, A. M. (2016). Study role of climate change in extreme threats to water quality. *Nature*, 535(7612), 349-350.
- Morris, D. P., & Lewis, W. M. (1988). Phytoplankton nutrient limitation in Colorado mountain lakes. *Freshwater Biology*, 20(3), 315–327. doi:10.1111/j.1365-2427.1988.tb00457.x
- Nöges, T. (2009). Relationships between morphometry, geographic location and water quality parameters of European lakes. *Hydrobiologia*, 633(1), 33–43. doi:10.1007/s10750-009-9874-x
- O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., et al. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, 42, 10,773–10,781. doi:10.1002/2015GL066235
- Paerl, H. W., Scott, J. T., McCarthy, M. J., Newell, S. E., Gardner, W. S., Havens, K. E., et al. (2016). It takes two to tango: When and where dual nutrient (N and P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science and Technology*, 50(20), 10805–10813. doi:10.1021/acs.est.6b02575
- Preston, D. L., Caine, N., McKnight, D. M., Williams, M. W., Hell, K., Miller, M. P., et al. (2016). Climate regulates alpine lake ice cover phenology and aquatic ecosystem structure. *Geophysical Research Letters*, 43(10), 5353–5360. doi:10.1002/2016GL069036
- Quinlan, R., A. Filazzola, A., Mahdian, O., Shuvo, A., Blagrove, K., Ewins, C., et al. (submitted). Relationships of total phosphorus and chlorophyll in lakes worldwide. *Limnology and Oceanography*.
- Rose, K. C., Greb, S. R., Diebel, M., & Turner, M. G., (2017). Annual precipitation regulates spatial and temporal drivers of lake water clarity: *Ecological Applications*, 27(2), 632–643.
- Sanchez-Lorenzo, A., Wild, M., Brunetti, M., Guijarro, J. A., Hakuba, M. Z., Calbó, J., Mystakidis, S., & Bartok, B. (2015). Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939–2012). *J. Geophys. Res. Atmos.*, 120, 9555–9569. doi:10.1002/2015JD023321.
- Schindler, D.W. (1997). Evolution of phosphorus limitation in lakes, *Science* 195, 260-262.
- Schmid, M., & Köster, O. (2016). Excess warming of a Central European lake driven by solar brightening. *Water Resources Research*, 52(10), 8103-8116.
- Schmid, M., Hunziker, S., & Wueest, A. (2014). Lake surface temperatures in a changing climate: a global sensitivity analysis. *Climatic Change*, 124, 301-315. doi:10.1007/s10584-014-1087.2
- Sharpley, A., Jarvie, H. P., Buda, A., May, L., Spears, B., & Kleinman, P. (2013). Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. *Journal of Environmental Quality*, 42(5), 1308–1326. doi:10.2134/jeq2013.03.0098
- Sinha, E., Michalak, A. M., & Balaji, V. (2017). Eutrophication will increase during the 21st century as a result of precipitation changes. *Science*, 357(6349), 1–5. doi:10.1126/science.aan2409
- Smith, V. H., & Schindler, D. W. (2009). Eutrophication science: where do we go from here? *Trends in Ecology and Evolution*, 24(4), 201–207. doi:10.1016/j.tree.2008.11.009
- Tian, W., Zhang, H., Zhao, L., Zhang, F., & Huang, H. (2017). Phytoplankton diversity effects on community biomass and stability along nutrient gradients in a eutrophic lake. *International Journal of Environmental Research and Public Health*, 14(1), 95. doi:10.3390/ijerph14010095

- Trolle, D., Hamilton, D. P., Pilditch, C. A., Duggan, I. C., & Jeppesen, E. (2011). Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management. *Environmental Modelling and Software*, 26(4), 354-370.
- Vaughan, I. P., & Gotelli, N. J. (2019). Water quality improvements offset the climatic debt for stream macroinvertebrates over twenty years. *Nature Communications*, 10(1), 1–8. doi:10.1038/s41467-019-09736-3
- Verburg, P., Hecky, R. E., & Kling, H. (2003). Ecological consequences of a century of warming in Lake Tanganyika. *Science*, 301(5632), 505-507.
- Webb, J. R., Hayes, N. M., Simpson, G. L., Leavitt, P. R., Baulch, H. M., & Finlay, K. (2019). Widespread nitrous oxide undersaturation in farm waterbodies creates an unexpected greenhouse gas sink. *Proceedings of the National Academy of Sciences of the United States of America*, 116(20), 9814–9819. doi:10.1073/pnas.1820389116
- Wild, M., Folini, D., Henschel, F., Fischer, N., & Müller, B. (2015). Projections of long-term changes in solar radiation based on CMIP5 climate models and their influence on energy yields of photovoltaic systems. *Solar Energy*, 116, 12–24. doi:10.1016/j.solener.2015.03.039
- Wild, M., Gilgen, H., Roesch, A., Ohmura, A., Long, C. N., Dutton, E. C., et al. (2005). From dimming to brightening: Decadal changes in solar radiation at earth's surface. *Science*, 308(5723), 847–850. doi:10.1126/science.1103215
- Wild, M. (2012). Enlightening global dimming and brightening. *Bulletin of the American Meteorological Society*, 93(1), 27–37. doi:10.1175/BAMS-D-11-00074.1
- Woelmer, W. M., Kao, Y. C., Bunnell, D. B., Deines, A. M., Bennion, D. H., Rogers, M. W., et al. (2016). Assessing the influence of watershed characteristics on chlorophyll a in waterbodies at global and regional scales. *Inland Waters*, 6(3), 379–392. doi:10.5268/IW-6.3.964
- Zhong, Y., Notaro, M., Vavrus, S. J., & Foster, M. J. (2016). Recent accelerated warming of the Laurentian Great Lakes: Physical drivers. *Limnology and Oceanography*, 61, 1762–1786. doi:10.1002/lno.10331

FIGURES

A)



B)

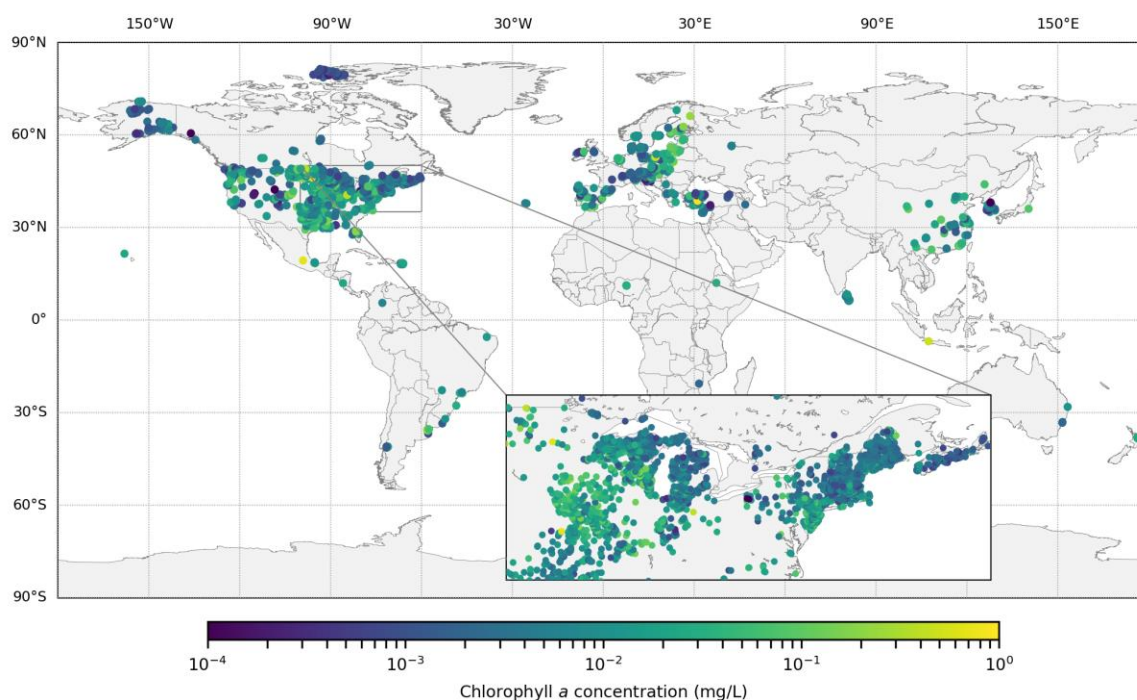


Figure 1. Maps showing lake locations for which water chemistry and climate data were collected for (a) 2561 lakes for which solar radiation data were available in addition to water chemistry and climate data (water temperature, precipitation, and cloud cover), (b) 3878 lakes for which there were only water chemistry and climate data available.

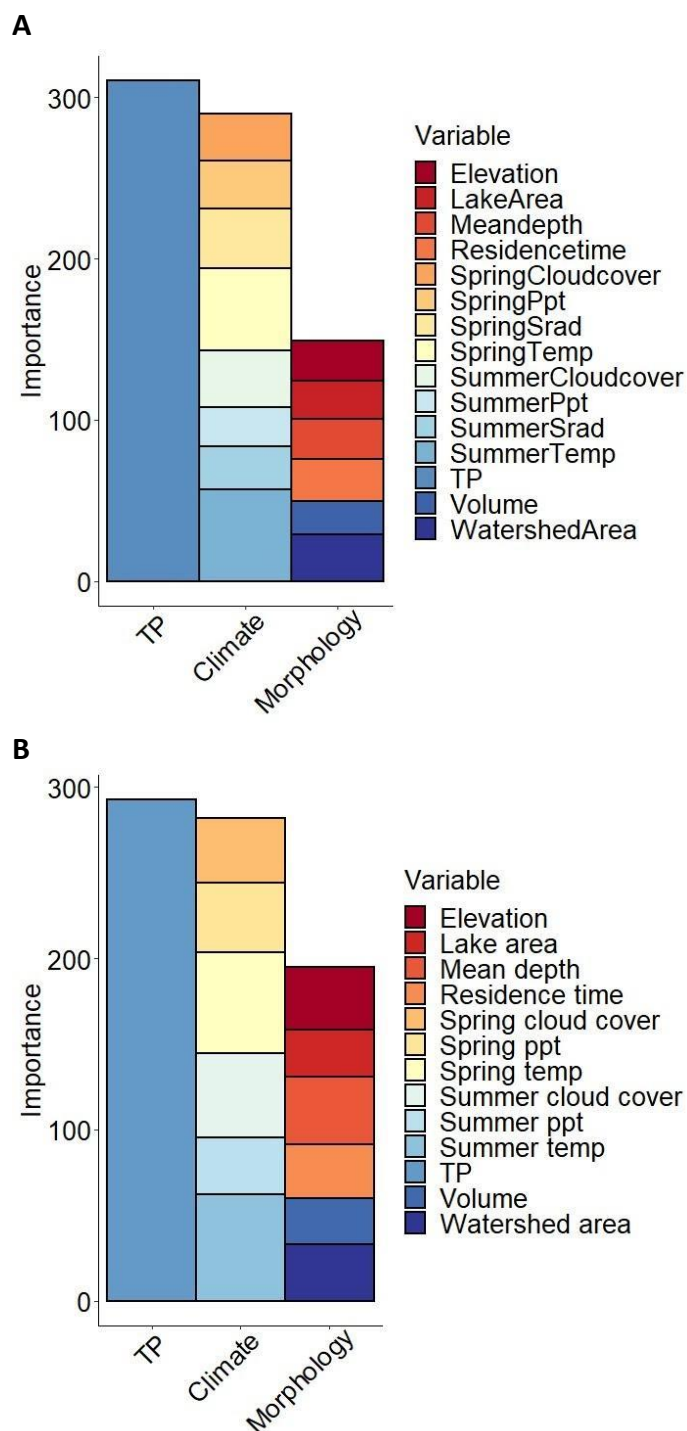


Figure 2. Stacked bar plot representing the proportion of variation explained by each variable in the random forest model. These variables are all drivers of chlorophyll in (A) the 2561-lake dataset which include solar radiation (D+SR) and (B) the 3878-lake dataset which does not include solar radiation data (D-SR). Variables are separated in groups characterized by nutrients (total phosphorus), climate (spring and summer solar radiation, cloud cover, precipitation, and air temperature), and morphology (elevation, lake area, mean depth, residence time, volume, watershed area).

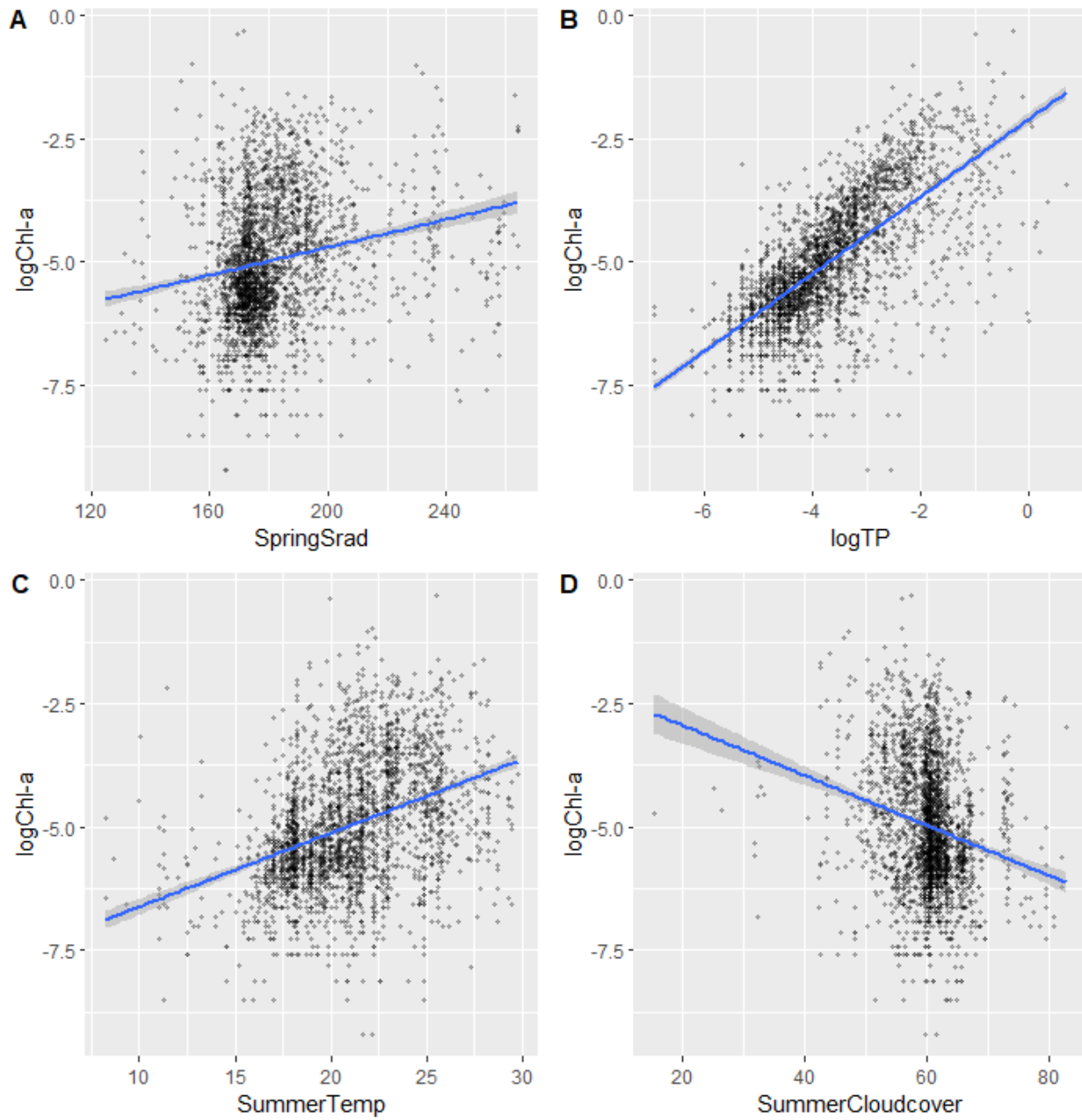


Figure 3. Chlorophyll concentrations plotted against the most influential climate and geomorphometric drivers for the 2561-lake dataset that includes solar radiation (D+SR), with linear models for (a) spring solar radiation ($R^2_{\text{adj}} = 0.041$), (b) total phosphorus ($R^2_{\text{adj}} = 0.47$), (c) summer air temperature ($R^2_{\text{adj}} = 0.14$), and (d) summer cloud cover ($R^2_{\text{adj}} = 0.045$).

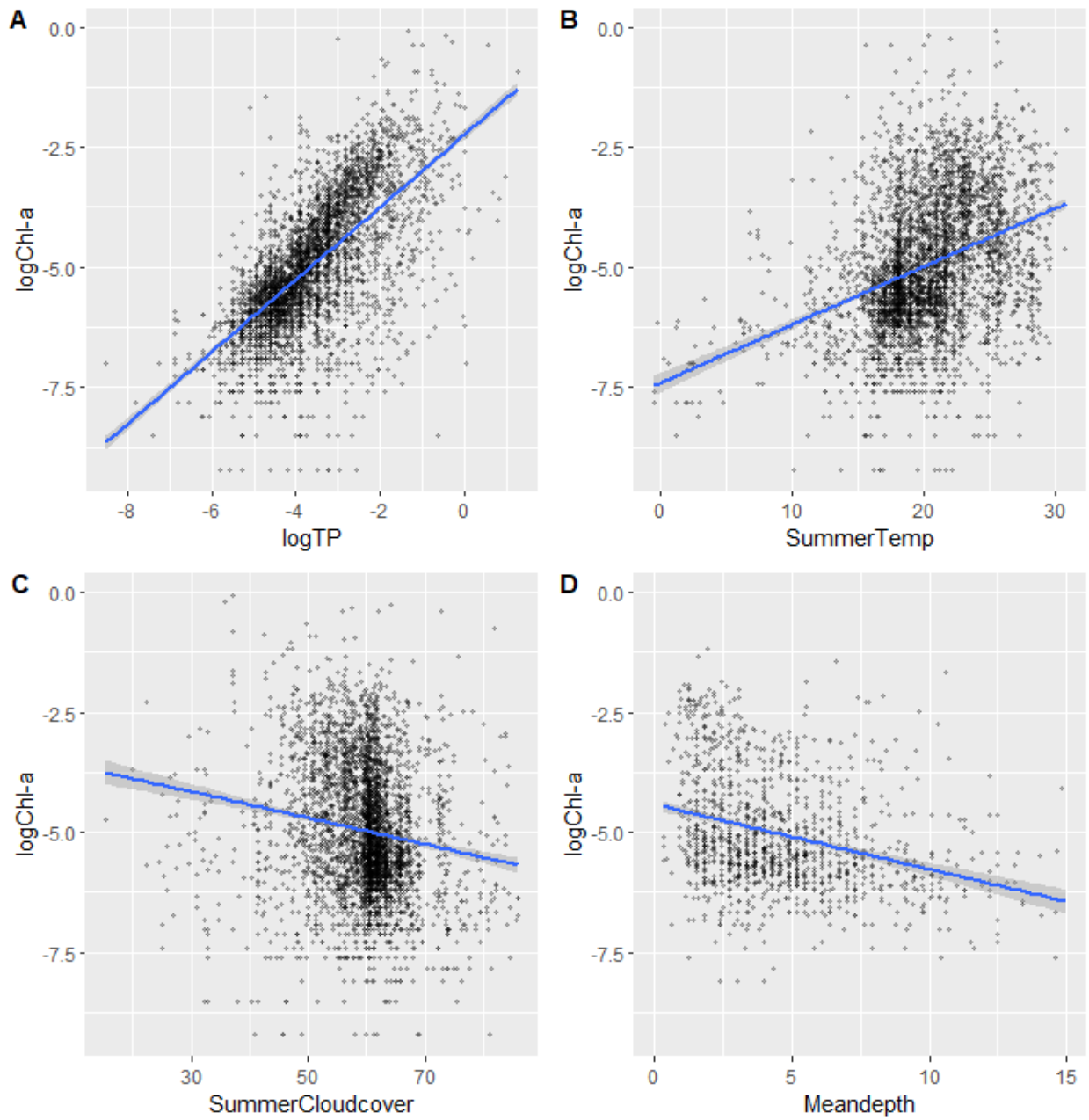
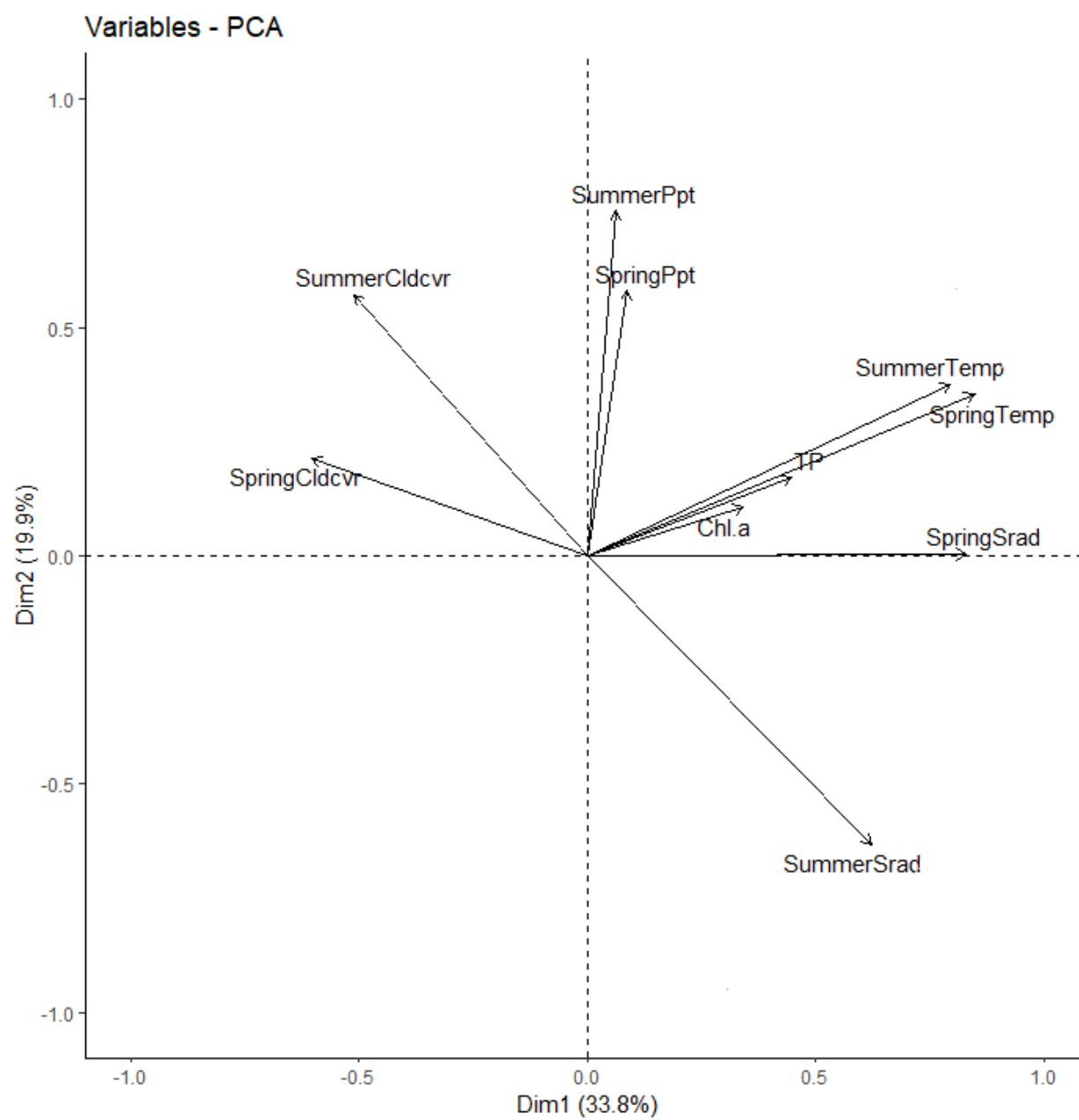


Figure 4. Chlorophyll concentrations plotted against the most influential drivers for the 3878-lake dataset without solar radiation (D-SR), with linear models for (a) total phosphorus concentrations ($R^2_{\text{adj}} = 0.0094$), (b) summer air temperature ($R^2_{\text{adj}} = 0.13$), (c) summer cloud cover ($R^2_{\text{adj}} = 0.023$) and (d), mean lake depth ($R^2_{\text{adj}} = 0.062$).

A



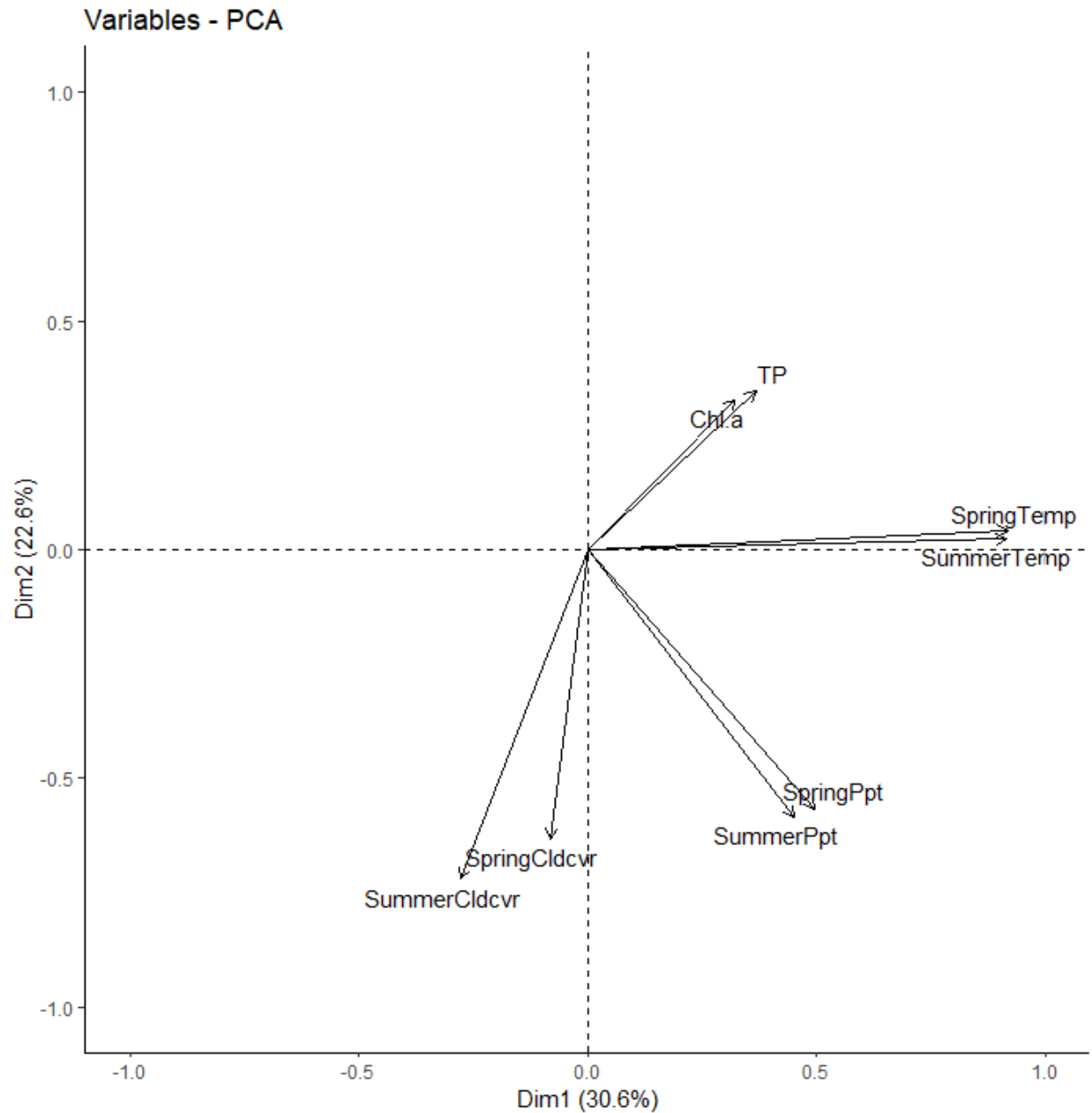
B

Figure 5. Principal Component Analysis for (a) the 2561-lake dataset (D+SR) and (b) the 387 8-lake dataset (D-SR). The most variation is explained by Principal Component 1 (Dim1), followed by the second principal component axis (Dim2). The arrows depict the environmental variables considered in the analysis. The longer the arrow, the more influential the variable. The angle between arrows suggests the correlation between variables. The acronyms are as follows: TP (total phosphorus), Temp (temperature), Ppt (precipitation), srad (solar radiation), Cldcvr (cloud cover)

SUPPLEMENTARY MATERIALS

Table S1. A summary table of the geomorphometric, water chemistry, and climate data extracted for both datasets encompassing 2561 (data including solar radiation data; (D+SR)) and 3878 lakes (dataset excluding solar radiation data (D-SR)). The units for each variable are listed next to the variable of interest. The mean and range has been calculated for each variable. *Data sourced from HydroLAKES (Messenger et al. 2016). Shoreline development measures the degree of deviation of a lake's surface shape from a circle; a value of 1 implies a perfectly circular shape and higher values indicate increased shoreline sinuosity (Messenger et al. 2016).

Variable	Units	Mean: D+SR	Range: D+ SR	Mean: D- SR	Range: D-SR
Year	---	2004	1982-2009	2003	1956-2018
Latitude	Decimal degrees	41.38	- 41.35 - 64.27	42.47	-41.35 – 81.33
Longitude	Decimal degrees	-79.65	-152.94 - 120.97	-68.84	-158.04 – 176.59
Elevation*	m	302.2	-1 - 3392	303.3	-1423
Area*	km ²	3.65	0.1 - 763.41	57.99	0.1 - 81843.92
Volume*	10 ⁶ m ³	10821019	111700-303068120	1.1E+07	111700 - 303068120
Mean depth*	m	4.4821	0.30 - 32.60	4.95	0.30 - 62.4
Residence time*	Days	629.8	0.1 - 131102.7	695.4	0.1 - 131102.7
Watershed area*	km ²	476.43	0.1 - 99952.60	1630.2	0.1 - 766137.4
Shoreline length*	10 ³ km	8.416	1.370-394.240	22.95	1.28 - 5248.36
Shore development*	---	10.534	1.280 - 696.8	2.007	1.040 - 21.160
Chla	mg/L	0.01643	0.00010 - 0.7300	0.01921	0.00010 - 0.93290
TP	mg/L	0.05593	0.001 - 1.99	0.06233	0.00020 - 3.6
Spring temperature	°C	9.025	-6.083 - 29.133	8.458	-56.917

Summer temperature	°C	20.975	8.283 - 29. 767	20.25	-31.22
Spring precipitation	mm	89.223	6.367 - 220.8	83.09	1.5 - 244.30
Summer precipitation	mm	91.92	0.1 - 262.73	90.79	0.1 - 384.03
Spring cloud cover	%	66.63	41.50 - 79.27	65.61	32.50 - 85.20
Summer cloud cover	%	60.19	15.43 - 82.93	59.4	15.43 - 85.93
Spring max radiation	$\text{kJ m}^{-2} \text{ day}^{-1}$	181.1	125.2 - 264.2	—	—
Summer max radiation	$\text{kJ m}^{-2} \text{ day}^{-1}$	235.2	165.3 - 326. 6	—	—

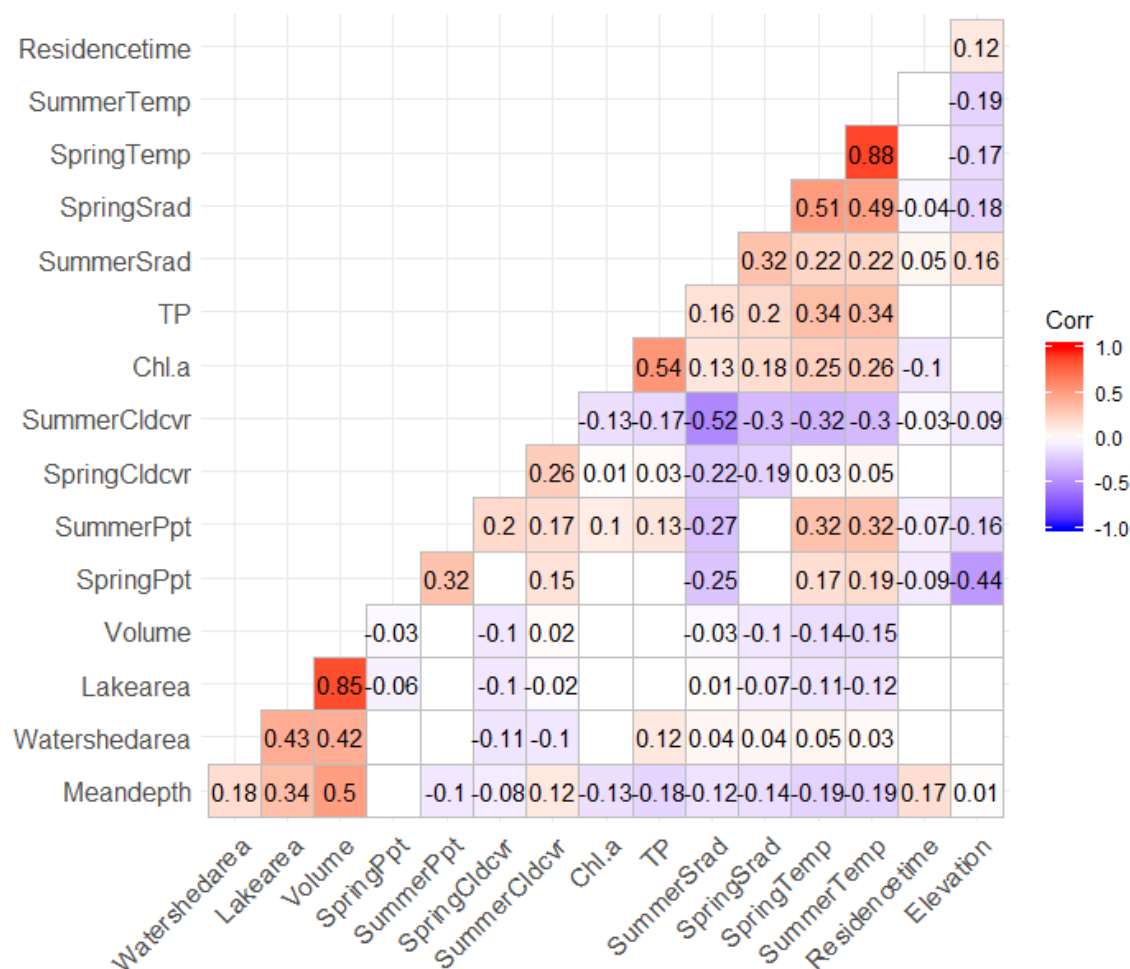


Figure S1. Kendall's Tau correlations between chlorophyll *a*, total phosphorus (TP), summer climate variables, spring climate variables, and geomorphometric characteristics for a) the dataset with solar radiation (D+SR). Warm colours indicate positive relationships, whereas cooler colours represent negative relationships. Squares which are left blank and white are done so because they are not significantly related ($P > 0.05$). The acronyms are as follows: TP (total phosphorus), Temp (temperature), Ppt (precipitation), srad (solar radiation), Cldcvr (cloud cover).

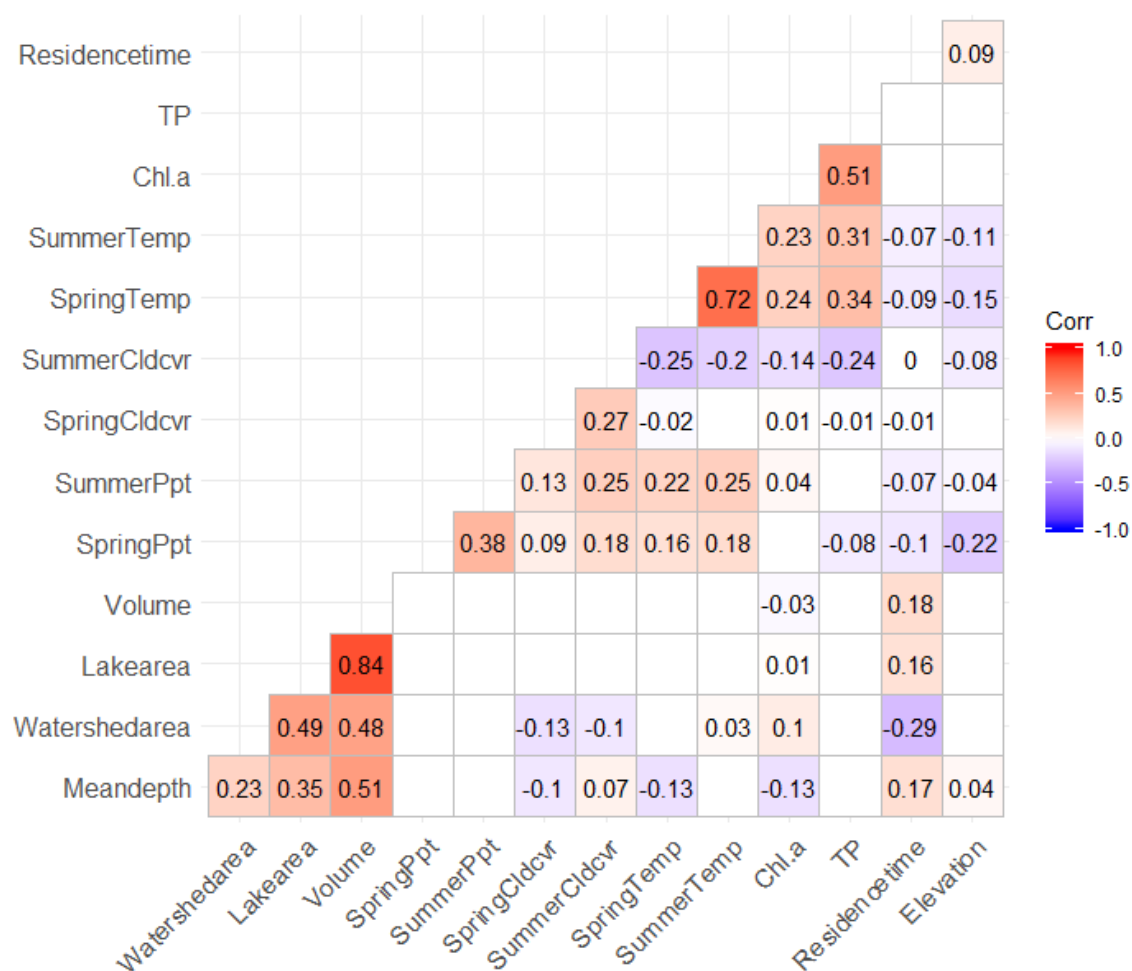


Figure S2. Kendall's Tau correlations between chlorophyll *a*, total phosphorus (TP), summer climate variables, spring climate variables, and geomorphometric characteristics for the larger dataset without solar radiation (D-SR). Warm colours indicate positive relationships, whereas cooler colours represent negative relationships. Squares which are left blank and white are done so because they are not significantly related ($P > 0.05$). The acronyms are as follows: TP (total phosphorus), Temp (temperature), Ppt (precipitation), srad (solar radiation), Cldcvr (cloud cover).

Chapter 2: Local, regional, and watershed spatial patterns structure chlorophyll *a* concentrations in lakes across the contiguous United States

Arnab Shuvo*¹, Derek Gray*², Luke Moslenko¹, Kevin Blaggrave¹, Carolyn Ewins¹,
Alessandro Filazzola¹, Octavia Mahdiyan¹, Catherine M. O'Reilly³, Roberto Quinlan¹,
and Sapna Sharma¹

¹Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada

²Department of Biology, Wilfrid Laurier University, Waterloo, Ontario, Canada

³Department of Geography, Geology, and the Environment, Illinois State University, Normal, Illinois, USA

***Joint first authors**

Keywords: Chlorophyll, Spatial Patterns, Phosphorus, Climate Change, Morphology, Land use, Lakes, US Water Quality

Chapter 2 will be submitted to Water.

Shuvo, A., Gray, D., Moslenko, L., Blaggrave, K., Ewins, C., Filazzola, A., Mahdiyan, O., O'Reilly, CM., Quinlan, R., Sharma, S. (Submitted). Local, regional, and watershed spatial patterns structure chlorophyll *a* concentrations in lakes across the contiguous United States. *Water*.

MY CONTRIBUTION

The manuscript titled, “Local, regional, and watershed spatial patterns structure chlorophyll *a* concentrations in lakes across the contiguous United States” will be submitted to the journal *Water*. The authors of this publication include Derek Gray, Luke Moslenko, Kevin Blagrove, Carolyn Ewins, Alessandro Filazzola, Octavia Mahdiyan, Catherine M. O’Reilly, Roberto Quinlan, and Sapna Sharma. I analyzed all the data, and wrote the Abstract, Introduction, Methods, Results, and Discussion. I also created the figures, and supplementary figures.

SUMMARY

Spatial patterns in water quality can be highly variable and the spatial variation within and among lakes can be structured at the regional (climate, atmospheric deposition, and regional landscape characteristics) and local spatial scales (lake morphometry, land use/ cover, geomorphic setting, and hydrology). Here we quantify the spatial patterns and drivers of chlorophyll *a* concentrations in 7194 lakes across the United States to answer: 1) What are the spatial patterns in water quality? 2) What are the relevant spatial signatures of chlorophyll *a* and its drivers? 3) What is the relative importance of nutrients, morphology, climate, and land use on chlorophyll? We synthesized a US wide database on lake water chemistry, with a focus on chlorophyll *a*, in addition to lake geomorphometry, climate, land use and human population for each lake. We found that together these variables explained 66% of the variation in Chl *a* concentrations. Local variables which include total phosphorus (TP) and lake morphology explained fifty percent of the unique and shared variation in chlorophyll levels. Spatial patterns explained an additional 23% of the variation. Sixteen percent of the variation in chlorophyll *a* was explained by climate, and 11% of the variation was explained by land use. Local variables, in particular TP, uniquely contributed 27% of the variation, whereas spatial patterns uniquely explained only 3% of the variation. Our research suggests that local spatial scales (distances <500km) and local factors (nutrients, and phytoplankton communities) are more influential than regional and national spatial scales for explaining primary production. As such, we recommend that management efforts and legislation should focus on drivers acting at local spatial scales, such as nutrient inputs, rather than only broader regional spatial scales.

INTRODUCTION

Spatial patterns in water quality can be highly variable, driven by many interacting factors (Fergus et al. 2016) including human population, nutrient inputs, climate, land use, and lake morphology (Peters & Meybeck, 2000; Liu et al., 2010; Saad et al., 2013; Green et al., 2015). The spatial variation within and among lakes can be found at the regional (climate, atmospheric deposition, and regional landscape characteristics) and local scales (lake morphometry, land use/ cover, geomorphic setting, and hydrology) (Soranno et al., 1999). Current trajectories in climate (Michalak, 2016), the continuing influence of nutrient management legacies (Sharpley et al., 2013), changes in land use associated with agriculture, industry, and recreation, and the growing demand for clean freshwater (UNEP), underscore the importance of understanding the relative importance of drivers of water quality. Furthermore, it is critical to identify interactions between various drivers and determine how these interactions vary spatially. These climate, land use, nutrient inputs, human population, and lake morphology drivers are influential in exacerbating water quality degradation through increases in primary production and algal biomass (de Oliveira Marcionilio et al., 2016). Here, we use chlorophyll *a* concentrations as a proxy of primary production as chlorophyll *a* is widely used to assess primary production owing to its ubiquity in a wide range of lake types and geophysical regions (de Oliveira Marcionilio et al., 2016).

Understanding the threat climate change poses to lake ecosystems around the globe (Rosenzweig et al., 2007; Michalak, 2016) is an urgent task for limnologists. Fortunately, climate-related signals and their responses are quite visible and easily measurable in lakes (Magnuson et al., 2000; Verburg et al., 2003; O'Reilly et al., 2015; Sharma et al., 2019). For example solar brightening and warming waters (O'Reilly et al., 2015), increased nutrients entering the lake

from precipitation events (Stockwell et al., 2020), and shifts in ice formation and thaw times (Magnuson et al., 2000) have all elicited responses individually and synergistically in phytoplankton communities (Hayes et al., 2015). Larger and more frequent storms can dislodge and mobilize large particles (Trenberth et al., 2003; Reichwaldt & Ghadouani 2012) wherein lakes may receive 80% - 400% of their average annual in-lake pools of N and P (Reichwaldt & Ghadouani 2012), which can be directly used by primary producers. Similarly, Collins et al. (2019) suggested that warmer summer air temperatures and increased winter precipitation have been directly linked to increased algal biomass in northeastern United States. Further, warmer temperatures and increased solar radiation inputs have contributed to earlier ice-out dates resulting in a longer growing season for phytoplankton communities and higher chlorophyll *a* levels (Woolway et al., 2020). In the presence of these climatic drivers there is a general increasing trend in algae and algal growth, however each lake has its own unique morphological characteristics which impact the degree to which algal growth may occur.

The morphometric features of a lake can directly or indirectly affect the limnological processes. Lake surface area may influence dilution capacity of the lake system (Moses et al., 2011), as found in Eastern China where lakes with smaller surface areas had greater chlorophyll *a* levels (Huang et al., 2014). However, larger catchment areas are susceptible to greater inputs of nutrient loads ultimately leading to increased primary production (Castello et al., 2016). Although drivers such as climate are largely influential in lake primary production, the degree and rate at which this occurs is dependent on the lake characteristics itself. Within lake processes such as residence time, sediment delivery, stratification, productivity and resuspension (Moses et al., 2011; Liu et al., 2012) in a lake system are largely influential in determining how much algae and algal growth can be sustained by the lake.

Further, land use can influence chlorophyll *a* concentrations. For example, lakes in the United States at the regional watershed scale and metropolitan area level have shown heterogeneity in hydrological response to urbanization (Poff et al., 2006; Brown et al., 2009; Hopkins et al., 2015) and agricultural land use (Morrice et al., 2008; Whaley & Luo, 2017). Hydrologic changes typically associated with urbanization include increased frequency and magnitude of high-flows, higher runoff efficiency, and reduced baseflow (Walsh et al., 2005). Studies conducted by Poff et al. (2006) and Brown et al. (2009) suggested urbanization increased runoff levels after large precipitation events in the southeastern and northwestern regions of the U.S., decreased in the southwestern U.S., and resulted in no change in the central region. Both studies concluded that there is great variability amongst different regions, and this may be attributed to the various usages of surrounding land (e.g. industry, increased populations, sewer systems). It is predicted that by 2030, 84% of the U.S. population is expected to reside in urban areas (United Nations Population Services, 2014) and such increases in impervious cover may alter hydrologic regimes. At the watershed scale, urbanization can cause shifts in runoff processes that directly impacts nutrient loading, carbon cycling, and the community composition and function of macroinvertebrates and fish (Hopkins et al., 2015). Similarly, watersheds surrounded by agricultural land use have shown high chlorophyll levels (Morrice et al., 2008; Wagner et al., 2011). Agricultural land use consists of croplands (farming practices) (Singh, 2002) and pasturelands (raising livestock) (Cournane et al., 2011). The application of pesticides for farming practices contributes high levels of potent phosphorus (Hu et al., 2013), while pasturelands are often subject to soil erosion events (Cournane et al., 2011). Both contribute to increased nutrient loading and higher chlorophyll levels in lakes (Wagner et al., 2011). Although lakes that are part of the same watershed share similar characteristics, local scale variations in

lakes need to be further understood to quantify the hydrologic changes that are driving declines in aquatic ecosystem health.

Our overall objective was to quantify the spatial patterns and drivers of chlorophyll *a* concentrations in 7194 lakes across the United States. More specifically, we asked: 1) What are the spatial patterns in water quality? 2) What are the relevant scales of spatial autocorrelation for chlorophyll *a* and its drivers? 3) What is the relative importance of nutrients, morphology, climate, and land use on chlorophyll variability among lakes? Based on previous studies, we hypothesized that total phosphorus would be the primary driver of variation in chlorophyll *a* levels among lakes. However, we also hypothesized that land use in lake catchments would be an important factor explaining variation in chlorophyll concentrations among lakes. Finally, we hypothesized that chlorophyll *a* levels and trophic status would show spatial structure in response to spatially structured environmental variables, but after factoring in the major spatially structured environmental variables, space would no longer be important.

METHODS

2.1 Data Acquisition and Dataset Compilation

We conducted a systematic review to develop a global chlorophyll database for approximately 11,000 lakes with over 250,000 sampling points (Filazzola et al., in review) (Figure 1). For this study, we used a subset of this database for the 48 contiguous states of the United States which included chlorophyll data for 7194 lakes. Each chlorophyll *a* measurement corresponded with the year in which the measurement was taken and the lake's latitude and longitude. In instances where the same lake was sampled in multiple locations, we took the median values of measured data for that lake. When a lake was sampled multiple times within the same year (e.g. monthly),

we extracted data from the time when chlorophyll *a* values were greatest. If a lake was sampled over multiple years, we selected data from the most recent year because it most closely resembles the most recent status of that lake.

Nutrients and morphology

Chemical and physical data were extracted during our systematic review when available, including: TP, total nitrogen (TN), dissolved organic carbon (DOC), dissolved oxygen (DO), pH, surface area, volume, maximum lake depth, mean lake depth, and Secchi depth (Filazzola et al., in review). Further, we used the HydroLAKES database (www.hydrosheds.org) to acquire morphological characteristics such as maximum and mean lake depth, elevation, surface area, water residence time, shoreline length, and watershed area for each lake greater than 10 ha in surface area (Messenger et al., 2016).

Climate

We extracted monthly air temperature (°C), precipitation (mm), cloud cover (%) for each lake, and sampling year from the University of East Anglia's Climate Research Unit (CRU) online open access database (Harris et al., 2020). CRU interpolates air temperature and precipitation from meteorological station measurements and summarizes the database on a grid at a 0.5° latitude and longitude resolution. Using the monthly mean temperature, precipitation, cloud cover and solar radiation values, we calculated the mean, maximum, and standard deviation annually, for each season, and at a lag of one year. Further, shortwave incoming solar radiation in the near infrared (0.2-4 μm) data were collected from the National Solar Radiation Database (NSRDB) between 1998 and 2014. For the scope of our study, we used spring (March, April, May) and summer (June, July, August) climate variables when algal growth is highest and to correspond to our summer chlorophyll values.

Land use and human population data

Land use data were collected from The Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) (Giglio et al., 2015). Version 6 provides global land cover types at yearly intervals (2001-2018). Land use data were extracted for 2015 to keep it consistent with the gridded population of the world (GPW) data whose most recent data file for 2015 (Center for International Earth Science Information Network - CIESIN - Columbia University, 2016). We calculated the proportion of 17 different landscape categories within the lake watershed (Giglio et al., 2015; Moslenko et al., in preparation) for each lake. Further, we acquired the distribution of human population (counts and densities) on a continuous global raster surface from the GPW at a resolution of 1km*1km (Center for International Earth Science Information Network - CIESIN - Columbia University, 2016).

2.2 Data analyses

We assessed normality of all variables using a Kolmogorov-Smirnov test and a qqplot using the “ks.test and qqnorm” functions in R. Chlorophyll *a* and total phosphorus (TP) were log-transformed to satisfy assumptions of normality. Spearman’s rank correlation coefficients were calculated between all variables using the “cor” function in R.

2.3 Spatial patterns and drivers of water quality

Spatial patterns and scales of water quality

We categorized all 7194 lakes by their trophic status using chlorophyll *a* based on the Trophic Status Index developed by Carlson (1977). Oligotrophic lakes were classified as those with chlorophyll *a* values below 0.0026 mg/L. Mesotrophic lakes were classified as lakes with chlorophyll *a* values between 0.0026 and 0.02 mg/L. Eutrophic lakes were categorized as lakes with chlorophyll *a* values between 0.02-0.056 mg/L. Lakes where chlorophyll values exceeded 0.056 were categorized as hyper-eutrophic (Carlson 1977). We used the Carlson (1977) index because it is currently the most cited index in the literature when classifying lake trophic statuses.

*Spatial signatures of chlorophyll *a* and its drivers*

We used Moran's *I* correlograms to identify the scale at which chlorophyll *a* and total phosphorus were spatially autocorrelated using the “*correlog*” function in the *pgirmess* package in R (Giradoux, 2018). The resulting correlogram shows the results of tests for spatial autocorrelation over a range of distance classes based on Moran's *I* (Moran, 1950). Moran's *I* values can range from -1 to 1, with values further from zero representing stronger positive or negative spatial autocorrelation (Moran, 1950). Positive (and significant) Moran's *I* values in a correlogram indicate that, for a given distance class, correlations within that class are more similar than is expected by chance; a negative (and significant) Mantel's *r* indicates the inverse, where correlations are less similar than is expected by chance.

To complement the Moran's *I* correlograms, we developed spatial predictor variables for our dataset using Moran Eigenvector Maps (MEMs; Dray et al., 2006). We constructed the MEMs using the *adespatial* package in R (Dray et al., 2012). Candidate spatial weighting matrices were

constructed with the `listw.candidates` function using a variety of connectivity matrices and weighting matrices. The connectivity matrices considered were Delaunay triangulation, minimum spanning tree, or nearest neighbour connection schemes, while the weighting matrices were based on linear, concave up, or concave down functions, or had no weights added (Borcard et al., 2011). The `listw.select` function was then used to test candidate spatial weighting matrices against the chlorophyll data. The `listw.select` function selects the best spatial weighting matrix and best subset of eigenvectors for use as spatial variables in models (Bauman et al., 2018). These spatial variables were then used in our analyses along with local physicochemical variables. We used a dataset of 2240 lakes for the MEM and variation partitioning analysis as there were no missing data in any of the predictor variables.

Relative importance of drivers of chlorophyll

We conducted a variation partitioning analysis to quantify the relative importance of spatial patterns using local variables, climate, land use, and space represented by X1, X2, X3, and X4 respectively (Borcard et al., 1992; Sharma et al., 2011). Each consisted of the following variables - local variables: TP, lake area, volume, average depth, residence time, elevation, watershed area; climate variables: summer temperature, spring temperature, spring solar radiation, summer solar radiation, summer precipitation, summer cloud cover, spring cloud cover; land use: croplands, woody savannas, deciduous broadleaf forests, mixed forests; and space: represented by the 88 significant MEMs.

In parallel, we conducted a random forest analysis to identify the nutrient (TP), climatic (spring and summer solar radiation, air temperature, precipitation, and cloud cover), and geomorphometric (lake mean depth, elevation, area, residence time, total volume, and watershed

area), land use, and human population drivers of chlorophyll *a* amongst the lakes. The primary variables of interest (Chl *a* and TP) were log₁₀ transformed before applying the randomforest function (package = randomforest) with iteration = 5 (Liaw and Wiener, 2002). Decision trees are the basic building blocks for a random forest. The decision tree is a machine learning algorithm which fits complex datasets with interacting predictor variables (Ali et al., 2012; Ben-Haim and Tom-Tov, 2010). Random forest analyses are classified as a strong learner because they are built as an ensemble of decision trees (Breiman, 2001; Ali et al., 2012). Random forest analyses work through bagging or bootstrap aggregation where subsets of the training data are randomly sampled, and a model is fit to these smaller data sets after which the predictions are aggregated (Ali et al., 2012; Breiman, 1996, 2001).

RESULTS

Spatial patterns and scales of water quality

The trophic status of more than half of American lakes are categorized as mesotrophic (moderate levels of Chl-*a*) and distributed across the United States (Figure 1). Over 25% of the lakes across the United States were categorized as eutrophic and hyper-eutrophic, suggesting that one quarter of American lakes were highly productive with degraded water quality. Hyper-eutrophic lakes tended to be found in the mid-west or in close proximity to large cities along the eastern shore (Figure 1). Eighteen percent of the lakes were categorized as oligotrophic and tended to be found in northern regions of the United States and along the Rocky Mountains (Figure 1).

*Spatial signatures of chlorophyll *a* and its drivers*

The spatial signatures of chlorophyll *a* and its drivers were classified based on distances at which they were spatially autocorrelated. Distances less than ~500 km were classified as local, 500-

1500 km as watershed, and >1500 km as regional scales. Spatial autocorrelation for chlorophyll (Figure 2a) was significant at distances less than ~1500 km suggesting that it has a spatial signature at the watershed level. Water chemistry variables, such as total phosphorus (Figure 2b), pH (Figure 2f), and morphological variables such as depth (Figure 2d) exhibited weak, but significant spatial autocorrelation at distances less than ~500 km, suggesting that they are spatially structured at the local scale. Elevation (Figure 2e) and proportion of croplands (Figure 2c) (Figure 2e) exhibited moderate spatial signatures at the watershed level. Climate variables such as summer precipitation (Figure 2g), solar radiation (Figure 2h), and temperature (Figure 2i) exhibited significant spatial autocorrelation at distances >1500 km suggesting that their spatial signatures are at the regional scale. Of the climate variables, the spatial signature for temperature and solar radiation are very broad and were found at distances >4000 km.

Spatial patterns in chlorophyll concentrations were significant at the regional, watershed, and local spatial scales. Moran Eigenvector Maps explained 36% of variation in chlorophyll without consideration of other drivers. The four MEMs that explained the largest amount of variation in chlorophyll levels among lakes are plotted in Figure 3. These four MEMs, 73, 364, 25, and 378, explained 2, 1.5, 1.4, and 1.3% of variation respectively (Figure 3). The intermediate spatial patterns (MEM 73 and 25) (Figure 3a, 3c) reflecting watershed spatial scales explained a combination of 3.4% of the variation, whereas the fine spatial patterns (MEM 364 and 378) (Figure 3b, 3d) representing local spatial scales explained 2.8% of the variation.

Relative importance of drivers of water quality

Variation partitioning analysis explained 66% of the variation in chlorophyll levels in freshwater lakes across the United States. 50% of the unique and shared variation in chlorophyll levels was

explained by local variables including TP and lake morphology (X1). Spatial patterns explained an additional 23% of the variation (X4). 16% of the variation in chlorophyll *a* was explained by climate (X2), and 11% of the variation was explained by land use (X3). Local variables, in particular TP, uniquely contributed 27% of the variation. After consideration of local, climate, and land-use variables, spatial patterns uniquely explained only 3% of the variation (Figure 4).

The random forest analysis on 2240 lakes dataset using nutrient concentrations, climatic variables, land use, and lake geomorphometric characteristics as predictors explained 61% of variation in lake chlorophyll levels in the United States. The random forest regression had a mean of squared residuals of 0.1109 suggesting that the model is a good fit (Supplementary figure 1). TP and the combination of climate variables explained similar amounts of variation in chlorophyll *a* (30% and 32% respectively). Summer and spring temperature followed by spring and summer solar radiation were the most influential of climate variables. Land use variables explained the second most variation (20%) where croplands were the most influential land use variable on lake chlorophyll production. Geomorphometric characteristics and human population explained the least variation (16% and 2%) respectively.

Of all the drivers assessed (Supplementary figure 2), TP was the most strongly correlated with Chl-*a* concentration, suggesting Chl-*a* increases with TP ($r = 0.59$, $P < 0.001$). Chl-*a* was positively correlated with spring and summer air temperature ($r = 0.28$, $P < 0.001$ and $r = 0.28$, $P < 0.001$), spring and summer solar radiation ($r = 0.16$, $P < 0.001$ and $r = 0.15$, $P < 0.001$), spring precipitation and summer precipitation ($r = 0.05$, $P < 0.001$ and $r = 0.14$, $P < 0.001$), and land use types: evergreen broadleaf forest ($r = 0.09$, $P < 0.001$), savannas ($r = 0.11$, $P < 0.001$), grasslands ($r = 0.04$, $P < 0.001$), permanent wetlands ($r = 0.04$, $P < 0.001$), croplands ($r = 0.33$, $P < 0.001$), permanent snow and ice ($r = 0.17$, $P < 0.001$), and cropland/natural vegetation mosaics ($r = 0.20$,

$P < 0.01$). Chlorophyll was negatively correlated with elevation ($r = -0.05$, $P < 0.001$), summer cloud cover ($r = -0.1$, $P < 0.001$), and land use types: evergreen needleleaf forests ($r = -0.15$, $P < 0.001$), deciduous broadleaf forests ($r = -0.24$, $P < 0.001$), mixed forests ($r = -0.29$, $P < 0.001$), open shrublands ($r = -0.04$, $P < 0.001$), woody savannas ($r = -0.22$, $P < 0.001$), permanent snow and ice ($r = -0.07$, $P < 0.001$), barren ($r = -0.03$, $P < 0.001$), and water bodies ($r = -0.12$, $P < 0.001$).

DISCUSSION

Our study quantified spatial patterns and their underlying drivers on chlorophyll *a* concentration in freshwater lakes across the United States. We found that the majority of lakes across the United States are mesotrophic (Figure 1). Furthermore, we showed that significant morphological and water quality drivers of chlorophyll, such as total phosphorus, lake depth, and pH, are spatially autocorrelated at the local scale ($< \sim 500\text{kms}$), whereas land use and altitudinal drivers of Chl *a* were spatially autocorrelated at the watershed scale ($\sim 500\text{-}1500\text{km}$). Lastly, climatic drivers, summer temperature, solar radiation, and precipitation, were spatially autocorrelated at the regional scale ($> 1500\text{km}$). Amongst the tested drivers, climate variables synergistically explained the most variation, however total phosphorus (TP) independently explained nearly as much. Land use, lake morphology, and human population were also important drivers in structuring lake chlorophyll *a* levels but explained significantly less variation compared to climate and TP. Spatial structure explained relatively little extra variation once nutrients, lake morphology, land use, and climate variables were considered.

Spatial patterns and scales of water quality

We found that over 25% of the lakes in the US are eutrophic or hyper eutrophic suggesting that they have high levels of chl-a. These lakes with degraded water quality are primarily situated in the midwest and northeast USA (Figure 1). Eutrophic lakes have high levels of biological productivity owing to the large inputs of nutrients they receive such as phosphorus (Chislock et al., 2013). Eutrophic lakes may initially have high levels of oxygen provided by a large number of plants but once a threshold is crossed the lake fauna suffers due to the high levels of respiration by the living vegetative matter (Chislock et al., 2013). This rate of change may be exacerbated by human interactions that lead to greater phosphorus loadings. Hypereutrophic lakes with even greater biological productivity and limited transparency (Kowalczywska-Madura et al., 2017) suffer from a very high supply of growth nutrients and have visibility limited to lower than 3 feet. There are potential dead zones where the overgrowth of algae often suffocates the fauna below the water depths and can no longer sustain any wildlife and/or plants (Kowalczywska-Madura et al., 2017). Mesotrophic lakes comprise the majority of US lakes with an intermediate level of productivity, although owing to their medium-level nutrients. Waters may be clear or brackish depending on nearby landscapes (Dittrich et al., 2013). Finally, oligotrophic lakes were found in the northern regions of the US and have the lowest proportion of the trophic classes (Tanabe et al., 2017). Oligotrophic lakes can support aquatic species who require well-oxygenated, cold waters, typically in colder regions of the world where mixing of nutrients is rare and slow due to the low water temperatures (Tanabe et al., 2017).

Spatial signatures of chlorophyll a and its drivers

We found that chlorophyll *a* concentrations were structured by spatial patterns and drivers acting at local, watershed, and regional scales using Moran's I correlograms and Moran Eigenvector Maps. For the scope of clarity and differentiation, based on the Moran's I correlograms we classified the local scale by limnological (chemical – nutrients (TP) and pH and lake morphometric (physical, i.e., mean depth) variables that influence water attributes, the watershed scale by landscape and land use characteristics, such as agricultural landscapes, and regional scales by climate variables (i.e., summer temperature, solar radiation, and precipitation) which operate at larger spatial scales (Carneiro et al., 2014; Wu et al., 2014; de Oliveira Marcionilio et al., 2015). The spatial signature range for summer temperature and solar radiation was exceptionally large and was positively spatially autocorrelated at distances >4000km similar to Di Cecco & Gouhier (2019) who suggest that the spatial signature of temperature may extend even further to the global spatial scale. These larger regional scales may act as constraints or drivers of the local factors (Ward et al., 1998; Miranda et al., 2014; de Oliveira Marcionilio et al., 2015). At the local scale nutrient availability, in particular TP, which is heavily linked to chlorophyll and primary production also exhibits positive spatial autocorrelation – which is fitting owing to the close ChlA – TP relationship (Fergus et al., 2016; Quinlan et al., in review).

Our results are supported by earlier studies conducted at smaller spatial scales which have suggested that lake chlorophyll levels are influenced by local, watershed, and regional drivers. For example, LaPierre et al (2015) suggested that ChlA was spatially autocorrelated for lakes situated 200km from one another in 7 different ecoregions in eastern Canada. Similarly, Soranno

et al. (2010), Wagner et al. (2011), and Filstrup et al. (2014) all worked with a dataset of 2105 lakes from 35 ecoregions collected from midwestern and northeastern United states. Here, strong ChlA and TP relations were identified but each varied by one order of magnitude (e.g. 3.5–69.4 ug/L and 7.6–168.1 ug/L respectively) between regions. It was suggested that management practices be conducted at smaller, more local scales because of such variation (Soranno et al., 2010; Wagner et al., 2011; Filstrup et al., 2014). Lakes adjacent to one another may share similar morphological characteristics owing to formation history, and likely belong to the same watershed with similar land use characteristics, such as agricultural (croplands) or urbanized lands. However, at the broad spatial scale, climate is a large overarching driver which influences chlorophyll levels as well as other drivers such as phosphorus

Relative importance of drivers of water quality

We found that TP was the single most important driver of chlorophyll based on correlational, variation partitioning, and random forest analyses. 50% of the variation in chlorophyll *a* levels was explained by local variables, consisting of TP and lake morphometry (X1 - Figure 4) and chlorophyll levels were highest in regions with high TP inputs ($r=0.59$) suggesting that local drivers are the most influential (Knowlton & Jones, 1997; Floury et al., 2003; Roozen et al., 2003; Gray et al., 2013; Carneiro et al., 2014; Miranda et al., 2014; Marcionilio et al., 2015). Nutrients, such as phosphorus, enter freshwater systems through diverse sources, including agriculture, urban stormwater runoff, and urban wastewater (Selman & Greenhalgh, 2010). Across the US, at the watershed scale, agriculture is the predominant contributor of nutrients through the usage of commercial fertilizers and animal manure which is rich in

phosphorus (Selman & Greenhalgh, 2010; Wagner et al., 2011; Ward et al., 2020). To that end, we found that many of the eutrophic and hypereutrophic lakes with elevated chlorophyll levels in our dataset were found in agricultural regions (spearman's correlation, $r=0.33$). It is anticipated that by 2050 the global population will reach > 9 billion people (UNPD, 2008), and with this growth the dietary demands are also expected to rise. To accommodate the needs of individuals, the demand for livestock production resulted in larger quantities of manure being produced by these operations and applied to land as fertilizer (Selman & Greenhalgh, 2010). Frequently, the rate and timing of land application of manure is dictated by the volume and availability of manure and not by crop needs. This results in ill-timed or overapplication of manure, further exacerbating nutrient runoff and leaching that increases lake nutrient loads and increased levels of primary production (Bosch et al., 2014).

The combination of climate variables, including summer temperature and solar radiation, explained the most amount of variation in the random forest model. Warmer water temperatures facilitate faster growth rates for summer algae in the northeastern region of the US (Collins et al., 2019), where we observed a high occurrence of eutrophic and hyper-eutrophic lakes. Over the span of 30 years (1980-2010) regions in North America have been experiencing solar brightening (Wild et al., 2012) and although these patterns vary spatially (Wild et al., 2012) and temporally (Sanchez-Lorenzo et al., 2015) many lake experiments have shown the influence of light limitation on algal growth (Kolzau et al., 2014; Dubourgh et al., 2015). In addition, increases in solar radiation have been linked to warming lake water temperatures globally (O'Reilly et al., 2015; Schmid and Köster, 2016; Zhong et al., 2016) and warmer waters have been associated with increases in chlorophyll and algal blooms (Ho et al., 2019). Interestingly,

once we accounted for drivers related to nutrients, lake morphometry, land use, human population, and climate, spatial patterns did not explain much unique variation confirming that we considered the predominant drivers of primary production in our analyses.

Our findings support recent literature which suggests that the local scales and local factors (nutrients, and phytoplankton communities) are more influential than regional and country-wide spatial scales for explaining primary production (Knowlton & Jones, 1997; Floury et al., 2003; Roozen et al., 2003; Gray et al., 2013; Carneiro et al., 2014; Miranda et al., 2014; Marcionilio et al., 2015). As mentioned previously TP has been identified as the key limiting factor for primary production at the local scale but this is influenced by regional factors such as land use, climate, and temporal variations (seasonality) (Canfield, 1983; Miranda et al., 2014; Marcionilio et al., 2015; Quinlan et al., in review). Aside from the influences of large-scale regional factors, there are interactions between local factors (TP and morphometry) as well that intensify the impacts and sensitivity of lakes to eutrophication. For example, regions with shallow lakes have higher TP than deeper lakes owing to greater internal loading (Sorrano et al., 2010). This may be intensified if the lakes or their respective watersheds are adjacent to agricultural land which have greater nutrient exports owing to their agricultural activities such as the use of fertilizers (Sorrano et al., 2010). As such, we recommend that management efforts and legislation should be proposed based on drivers acting at local spatial scales rather than large regional spatial scales.

ACKNOWLEDGEMENTS

Data are available in Filazzola et al. (in revision). We thank the numerous data providers who provided data for this study. Funding for this study was provided by the Ontario Ministry of Innovation Early Researcher Award, York Research Chair in Global Change Biology, and Natural Sciences Engineering Research Council Discovery Grant to SS, and the Ontario Ministry of Environment and Climate Change Best in Science Grant to DG and SS. SS and DG secured funding for the project; SS led the project. AF, KB, AS, LM, CE, and OM assembled the dataset; AS, AF, and DG conducted analyses; KB and AS made figures; SS, COR, AS, OM, and DG wrote sections of the text. All authors contributed ideas, editing and revision. Authors between the co-leads and last author have similar contributions and are listed in alphabetical order.

REFERENCES

- Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *International Journal of Computer Science Issues*, 9(5), 272–278.
- Bauman, D., Drouet, T., Fortin, M. J., & Dray, S. (2018). Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. *Ecology*, 99(10), 2159–2166.
- Ben-Haim, Y., & Tom-Tov, E. (2010). A streaming parallel decision tree algorithm. *Journal of Machine Learning Research*, 11, 849–872
- Bosch, N. S., Evans, M. A., Scavia, D., & Allan, J. D. (2014). Interacting effects of climate change and agricultural BMPs on nutrient runoff entering Lake Erie. *Journal of Great Lakes Research*, 40(3), 581–589.
- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(421), 123–140. doi:10.1007/BF00058655
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. doi:10.1201/9780429469275-8
- Brown, L.R., Cuffney, T.F., Coles, J.F., Fitzpatrick, F., McMahon, G., Steuer, J., Bell, A.H. and May, J.T. (2009). Urban streams across the USA: lessons learned from studies in 9 metropolitan areas. *Journal of the North American Benthological Society*, 28(4), 1051–1069.
- Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73(3), 1045–1055.
- Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology with R*.
- Canfield Jr, D. E., Bachmann, R. W., & Hoyer, M. V. (1983). Freeze-out of salts in hard-water lakes 1. *Limnology and Oceanography*, 28(5), 970–977.
- Carlson, R. E. (1977). A trophic state index for lakes 1. *Limnology and oceanography*, 22(2), 361–369.
- Carneiro, F. M., Nabout, J. C., Vieira, L. C., Roland, F., & Bini, L. M. (2014). Determinants of chlorophyll-a concentration in tropical reservoirs. *Hydrobiologia*, 740(1), 89–99.
- Castello, L., & Macedo, M. N. (2016). Large-scale degradation of Amazonian freshwater ecosystems. *Global change biology*, 22(3), 990–1007.
- Chislock, M. F., Doster, E., Zitomer, R. A., & Wilson, A. E. (2013). Eutrophication: causes, consequences, and controls in aquatic ecosystems. *Nature Education Knowledge*, 4(4), 10.
- CIESIN, S. (2016). Gridded population of the world, version 4 (GPWV4): population density. Technical report, Center for International Earth Science Information Network-CIESIN-Columbia University. NASA Socioeconomic Data and Applications Center (SEDAC).
- Collins, S.M., Yuan, S., Tan, P.N., Oliver, S.K., Lapierre, J.F., Cheruvelil, K.S., Fergus, C.E., Skaff, N.K., Stachelek, J., Wagner, T. and Soranno, P.A. (2019). Winter precipitation and summer temperature predict lake water quality at macroscales. *Water Resources Research*, 55(4), 2708–2721.
- Cournane, F. C., McDowell, R., Littlejohn, R., & Condrón, L. (2011). Effects of cattle, sheep and deer grazing on soil physical quality and losses of phosphorus and suspended sediment losses in surface runoff. *Agriculture, ecosystems & environment*, 140(1–2), 264–272.
- de Oliveira Marcionilio, S. M. L., Alves, M. T. R., Borges, P. P., Machado, K. B., Araújo, C. S. T., da Cunha, H. F., & Nabout, J. C. (2015). The state of global scientific literature on chlorophyll-A. *Bioscience Journal*, 31(3).
- de Oliveira Marcionilio, S.M.L., Machado, K.B., Carneiro, F.M., Ferreira, M.E., Carvalho, P., Vieira, L.C.G., de Moraes Huszar, V.L. and Nabout, J.C. (2016). Environmental factors affecting

- chlorophyll-a concentration in tropical floodplain lakes, Central Brazil. *Environmental monitoring and assessment*, 188(11), 611.
- Di Cecco, G. J., & Gouhier, T. C. (2018). Increased spatial and temporal autocorrelation of temperature under climate change. *Scientific reports*, 8(1), 1-9.
- Dittrich, M., Chesnyuk, A., Gudimov, A., McCulloch, J., Quazi, S., Young, J., Winter, J., Stainsby, E. and Arhonditsis, G. (2013). Phosphorus retention in a mesotrophic lake under transient loading conditions: insights from a sediment phosphorus binding form study. *water research*, 47(3), 1433-1447.
- Dubourg, P., North, R. L., Hunter, K., Vandergucht, D. M., Abirhire, O., Silsbe, G. M., et al. (2015). Light and nutrient co-limitation of phytoplankton communities in a large reservoir: Lake Diefenbaker, Saskatchewan, Canada. *Journal of Great Lakes Research*, 41, 129–143. doi: 10.1016/j.jglr.2015.10.001
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *ecological modelling*, 196(3-4), 483-493.
- Dray, S., Péliissier, R., Couteron, P., Fortin, M.J., Legendre, P., Peres-Neto, P.R., Bellier, E., Bivand, R., Blanchet, F.G., De Cáceres, M. and Dufour, A.B. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, 82(3), 257-275.
- Fergus, C. E., Finley, A. O., Soranno, P. A., & Wagner, T. (2016). Spatial variation in nutrient and watercolor effects on lake chlorophyll at macroscales. *Plos one*, 11(10), e0164592.
- Filazzola, A., Mahdiyan O., Shuvo, A., Ewins, C., Moslenko, L., Sadid, T., Blaggrave, K., Gray, D., Quinlan, R., O'Reilly, CM., Sharma, S. (In review). A global database of chlorophyll and water chemistry in freshwater lakes. *Scientific Data*.
- Filstrup, C. T., Wagner, T., Soranno, P. A., Stanley, E. H., Stow, C. A., Webster, K. E., & Downing, J. A. (2014). Regional variability among nonlinear chlorophyll—phosphorus relationships in lakes. *Limnology and Oceanography*, 59(5), 1691-1703.
- Giglio, L., Schroeder, W., Hall, J. V., & Justice, C. O. (2015). MODIS collection 6 active fire product user's guide revision A. Department of Geographical Sciences. University of Maryland.
- Gray, B. R., Rogala, J. R., & Houser, J. N. (2013). Treating floodplain lakes of large rivers as study units for variables that vary within lakes; an evaluation using chlorophyll a and inorganic suspended solids data from floodplain lakes of the upper mississippi river. *River Research and Applications*, 29(3), 330-342.
- Hayes, N. M., Vanni, M. J., Horgan, M. J., & Renwick, W. H. (2015). Climate and land use interactively affect lake phytoplankton nutrient limitation status. *Ecology*, 96(2), 392-402.
- Ho, J.C., Michalak, A.M., & Pahlevan, N., (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, 574, 667–670. doi:10.1038/s41586-019-1648-7.
- Hopkins, K.G., Morse, N.B., Bain, D.J., Bettez, N.D., Grimm, N.B., Morse, J.L., Palta, M.M., Shuster, W.D., Bratt, A.R. and Suchy, A.K. (2015). Assessment of regional variation in streamflow responses to urbanization and the persistence of physiography. *Environmental science & technology*, 49(5), 2724-2732.
- Huang, J., Xu, Q., Xi, B., Wang, X., Jia, K., Huo, S., Su, J., Zhang, T. and Li, C. (2014). Effects of lake-basin morphological and hydrological characteristics on the eutrophication of shallow lakes in eastern China. *Journal of Great Lakes Research*, 40(3), 666-674.
- Hu, X., Roberts, D.P., Xie, L., Maul, J.E., Yu, C., Li, Y., Zhang, S. and Liao, X. (2013). Development of a biologically based fertilizer, incorporating *Bacillus megaterium* A6, for

- improved phosphorus nutrition of oilseed rape. *Canadian journal of microbiology*, 59(4), 231-236.
- Kolzau, S., Wiedner, C., Rücker, J., Köhler, J., Köhler, A., & Dolman, A. M. (2014). Seasonal patterns of Nitrogen and Phosphorus limitation in four German Lakes and the predictability of limitation status from ambient nutrient concentrations. *PLoS ONE*, 9(4). doi: 10.1371/journal.pone.0096065.
- Kowalczywska-Madura, K., Dondajewska, R., Gołdyn, R., & Podsiadłowski, S. (2017). The influence of restoration measures on phosphorus internal loading from the sediments of a hypereutrophic lake. *Environmental Science and Pollution Research*, 24(16), 14417-14429.
- Knowlton, M. F., & Jones, J. R. (1997). Trophic status of Missouri River floodplain lakes in relation to basin type and connectivity. *Wetlands*, 17(4), 468-475.
- 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.
- Liaw, A., & M. Wiener. (2002). Classification and Regression by randomForest. *R News* 2/3: 18–22.
- Liu, W., Zhang, Q., & Liu, G. (2010). Lake eutrophication associated with geographic location, lake morphology and climate in China. *Hydrobiologia*, 644(1), 289-299.
- Liu, W., Li, S., Bu, H., Zhang, Q., & Liu, G. (2012). Eutrophication in the Yunnan Plateau lakes: the influence of lake morphology, watershed land use, and socioeconomic factors. *Environmental Science and Pollution Research*, 19(3), 858-870.
- Magnuson, J. J., Wynne, R. H., Benson, B. J., & Robertson, D. M. (2000). Lake and river ice as a powerful indicator of past and present climates. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen*, 27(5), 2749-2756.
- Messenger, 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), 1-11.
- Michalak, A. M. (2016). Study role of climate change in extreme threats to water quality. *Nature*, 535(7612), 349-350.
- Miranda, L. E., Andrews, C. S., & Kröger, R. (2014). Connectedness of land use, nutrients, primary production, and fish assemblages in oxbow lakes. *Aquatic sciences*, 76(1), 41-50.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23.
- Moses, S. A., Janaki, L., Joseph, S., Justus, J., & Vimala, S. R. (2011). Influence of lake morphology on water quality. *Environmental monitoring and assessment*, 182(1-4), 443-454.
- O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J., Schneider, P., Lenters, J.D., McIntyre, P.B., Kraemer, B.M. and Weyhenmeyer, G.A. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, 42(24), 10-773.
- Peters, N. E., & Meybeck, M. (2000). Water quality degradation effects on freshwater availability: impacts of human activities. *Water International*, 25(2), 185-193.
- Poff, N. L., Bledsoe, B. P., & Cuhaciyan, C. O. (2006). Hydrologic variation with land use across the contiguous United States: geomorphic and ecological consequences for stream ecosystems. *Geomorphology*, 79(3-4), 264-285.
- Quinlan, R., A. Filazzola, A., Mahdian, O., Shuvo, A., Blagrove, K., Ewins, C., et al. (submitted). Relationships of total phosphorus and chlorophyll in lakes worldwide. *Limnology and Oceanography*.

- Reichwaldt, E. S., & Ghadouani, A. (2012). Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. *Water research*, 46(5), 1372-1393.
- Roozen, F. C., Van Geest, G. J., Ibelings, B. W., Roijackers, R., Scheffer, M., & Buijse, A. D. (2003). Lake age and water level affect the turbidity of floodplain lakes along the lower Rhine. *Freshwater Biology*, 48(3), 519-531.
- Rosenzweig, C., Casassa, G., Karoly, D.J., Imeson, A., Liu, C., Menzel, A., Rawlins, S., Root, T.L., Seguin, B., Tryjanowski, P. and Parry, M.L. (2007). Assessment of observed changes and responses in natural and managed systems.
- Sanchez-Lorenzo, A., Wild, M., Brunetti, M., Guijarro, J. A., Hakuba, M. Z., Calbó, J., Mystakidis, S., & Bartok, B. (2015). Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939–2012). *J. Geophys. Res. Atmos.*, 120, 9555– 9569. doi:10.1002/2015JD023321.
- Schmid, M., & Köster, O. (2016). Excess warming of a Central European lake driven by solar brightening. *Water Resources Research*, 52(10), 8103-8116.
- Selman, M., & Greenhalgh, S. (2010). Eutrophication: sources and drivers of nutrient pollution. *Renewable Resources Journal*, 26(4), 19-26.
- Sharma, S., Legendre, P., De Cáceres, M., & Boisclair, D. (2011). The role of environmental and spatial processes in structuring native and non-native fish communities across thousands of lakes. *Ecography*, 34(5), 762-771.
- Sharma, S., Blagrove, K., Magnuson, J.J., O'Reilly, C.M., Oliver, S., Batt, R.D., Magee, M.R., Straile, D., Weyhenmeyer, G.A., Winslow, L. and Woolway, R.I. (2019). Widespread loss of lake ice around the Northern Hemisphere in a warming world. *Nature Climate Change*, 9(3), 227-231.
- Sharpley, A., Jarvie, H. P., Buda, A., May, L., Spears, B., & Kleinman, P. (2013). Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. *Journal of environmental quality*, 42(5), 1308-1326.
- Singh, S. P. (2002). Balancing the approaches of environmental conservation by considering ecosystem services as well as biodiversity. *Current Science*, 82(11), 1331-1335.
- Soranno, P.A., Webster, K.E., Riera, J.L., Kratz, T.K., Baron, J.S., Bukaveckas, P.A., Kling, G.W., White, D.S., Caine, N., Lathrop, R.C. and Leavitt, P.R. (1999). Spatial variation among lakes within landscapes: ecological organization along lake chains. *Ecosystems*, 2(5), 395-410.
- Soranno, P. A., Cheruvilil, K. S., Webster, K. E., Bremigan, M. T., Wagner, T., & Stow, C. A. (2010). Using landscape limnology to classify freshwater ecosystems for multi-ecosystem management and conservation. *BioScience*, 60(6), 440-454.
- Stockwell, J.D., Doubek, J.P., Adrian, R., Anneville, O., Carey, C.C., Carvalho, L., De Senerpont Domis, L.N., Dur, G., Frassl, M.A., Grossart, H.P. and Ibelings, B.W. (2020). Storm impacts on phytoplankton community dynamics in lakes. *Global change biology*, 26(5), 2756-2784.
- Tanabe, Y., Yasui, S., Osono, T., Uchida, M., Kudoh, S., & Yamamuro, M. (2017). Abundant deposits of nutrients inside lakebeds of Antarctic oligotrophic lakes. *Polar Biology*, 40(3), 603-613.
- Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, 84(9), 1205-1218.
- Unpd. (2008). World population prospects: the 2008 revision population database.
- Verburg, P., Hecky, R. E., & Kling, H. (2003). Ecological consequences of a century of warming in Lake Tanganyika. *Science*, 301(5632), 505-507.

- Wagner, T., Soranno, P. A., Webster, K. E., & Cheruvilil, K. S. (2011). Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes. *Freshwater Biology*, 56(9), 1811-1824.
- Walsh, C. J., Roy, A. H., Feminella, J. W., Cottingham, P. D., Groffman, P. M., & Morgan, R. P. (2005). The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society*, 24(3), 706-723.
- Ward, J. V., Bretschko, G., Brunke, M., Danielopol, D., Gibert, J., Gonser, T., & Hildrew, A. G. (1998). The boundaries of river systems: the metazoan perspective. *Freshwater Biology*, 40(3), 531-569.
- Ward, N. K., Steele, B. G., Weathers, K. C., Cottingham, K. L., Ewing, H. A., Hanson, P. C., & Carey, C. C. (2020). Differential responses of maximum versus median chlorophyll-a to air temperature and nutrient loads in an oligotrophic lake over 31 years. *Water Resources Research*, 56(7), e2020WR027296.
- Wild, M. (2012). Enlightening global dimming and brightening. *Bulletin of the American Meteorological Society*, 93(1), 27–37. doi:10.1175/BAMS-D-11-00074.1
- Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., & Sharma, S. (2020). Global lake responses to climate change. *Nature Reviews Earth & Environment*, 1-16.
- Wu, Z., He, H., Cai, Y., Zhang, L., & Chen, Y. (2014). Spatial distribution of chlorophyll a and its relationship with the environment during summer in Lake Poyang: a Yangtze-connected lake. *Hydrobiologia*, 732(1), 61-70.
- Zhong, Y., Notaro, M., Vavrus, S. J., & Foster, M. J. (2016). Recent accelerated warming of the Laurentian Great Lakes: Physical drivers. *Limnology and Oceanography*, 61, 1762–1786. doi:10.1002/lno.10331

FIGURES

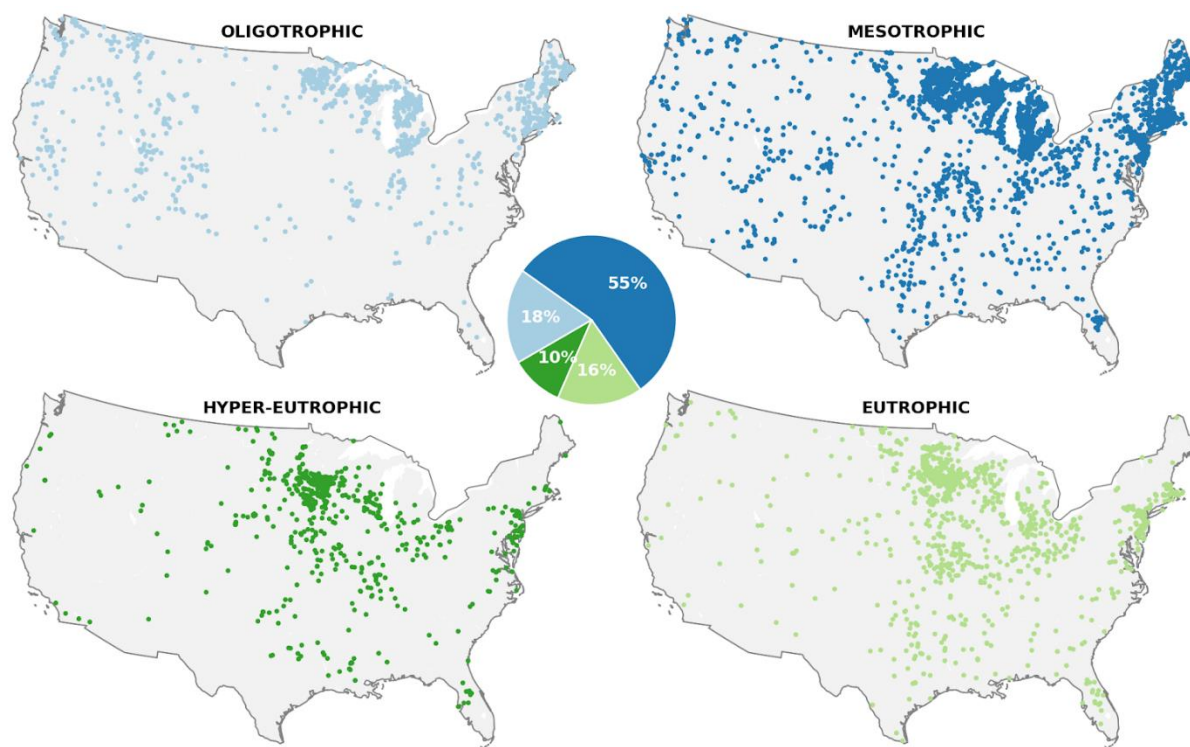


Figure 1. Maps showing lake locations for which water chemistry and climate data were collected in the United States. Lake trophic status is indicated by colour based on the Carlson trophic index Chl-*a* thresholds, where measures below 0.0026 mg/L are considered oligotrophic, measures between 0.0026 and 0.020 mg/L are mesotrophic, between 0.020 and 0.056 mg/L are considered eutrophic, and above 0.056 are hyper-eutrophic.

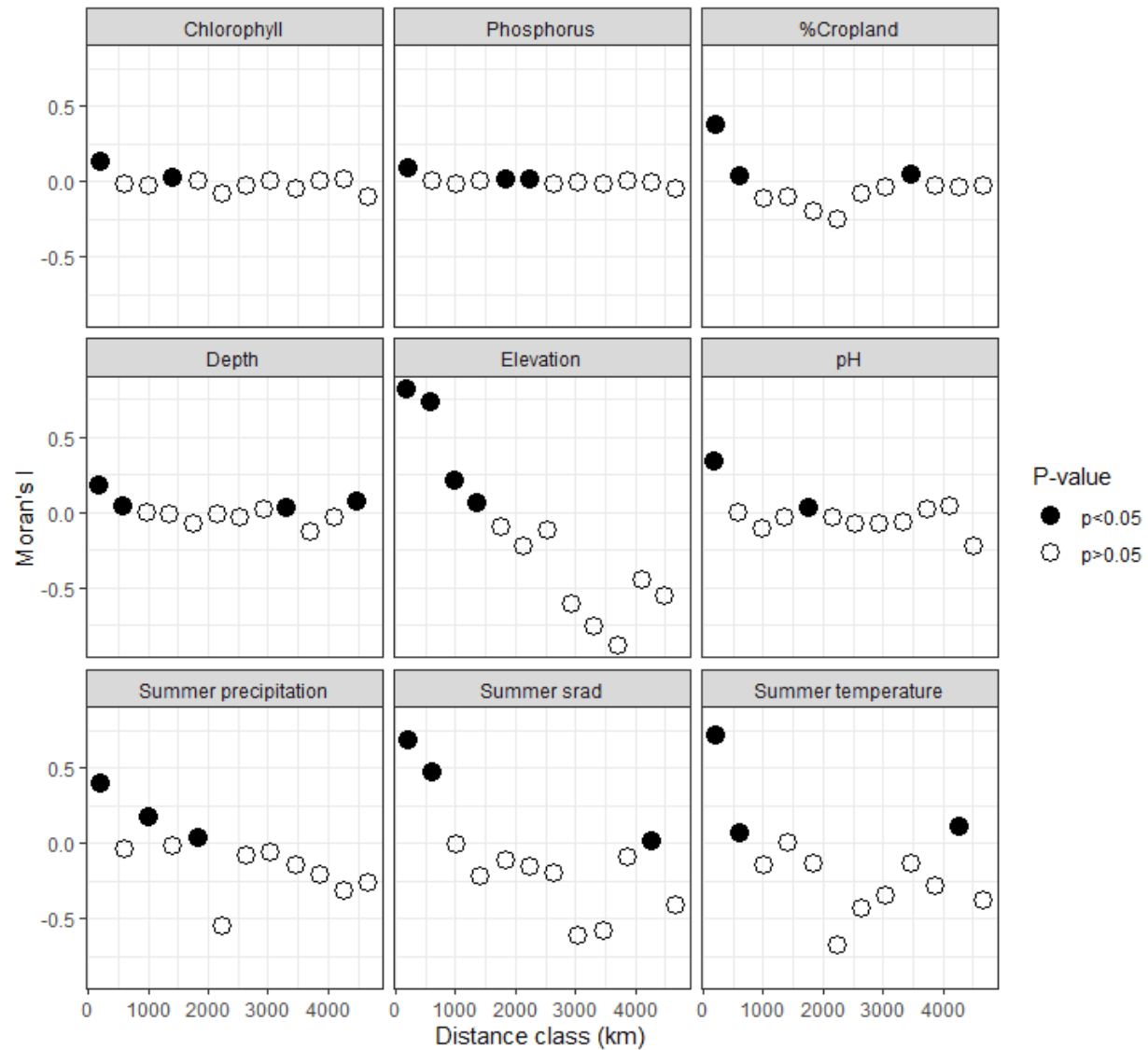


Figure 2. Plots of Moran's I showing spatial autocorrelation in chlorophyll, phosphorus values, percentage cropland, mean depth, elevation, pH, summer precipitation, summer solar radiation, and summer temperature values ranging from 0-4000kms. Filled circles denote significant spatial autocorrelation at $p < 0.05$.

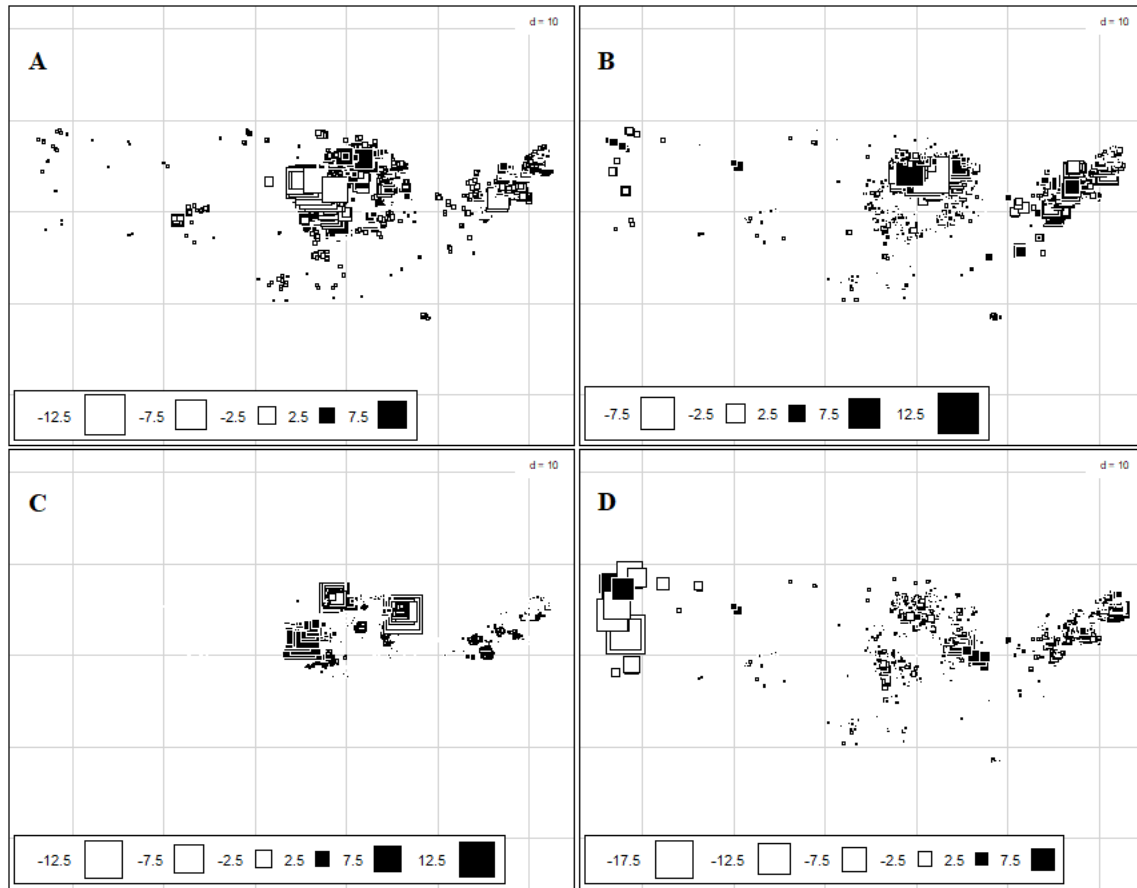


Figure 3. Bubble plots highlighting the four MEM's that explained the highest variation in Chl-a. The combination of all significant MEM variables explained 36% of the variation in Chl-a. MEM 73 (a) and MEM25 (c) explain 3.4% of the variation and reflect spatial patterns acting at intermediate and watershed level scales. MEM364 (b) and MEM378 (d) explain 2.8% of the variation and represent fine spatial patterns acting at local spatial scales.

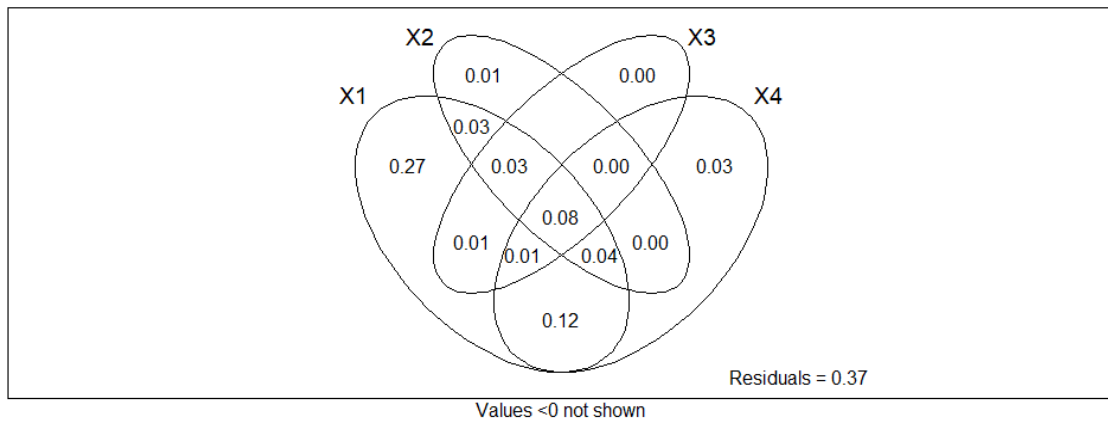


Figure 4. Variation partitioning analysis representing the four major groups and their relative influence on chlorophyll production in freshwater lakes across the United States. X1 is representative of local variables: TP, lake area, volume, average depth, residence time, elevation, watershed area. X2 represents climate: summer temperature, spring temperature, spring solar radiation, summer solar radiation, summer precipitation, summer cloud cover, and spring cloud cover. X3 represents the land use variables and human population: croplands, woody savannas, deciduous broadleaf forests, mixed forests, and population. X4 highlights the spatial structure through the 88 significant MEMs.

SUPPLEMENTARY MATERIALS

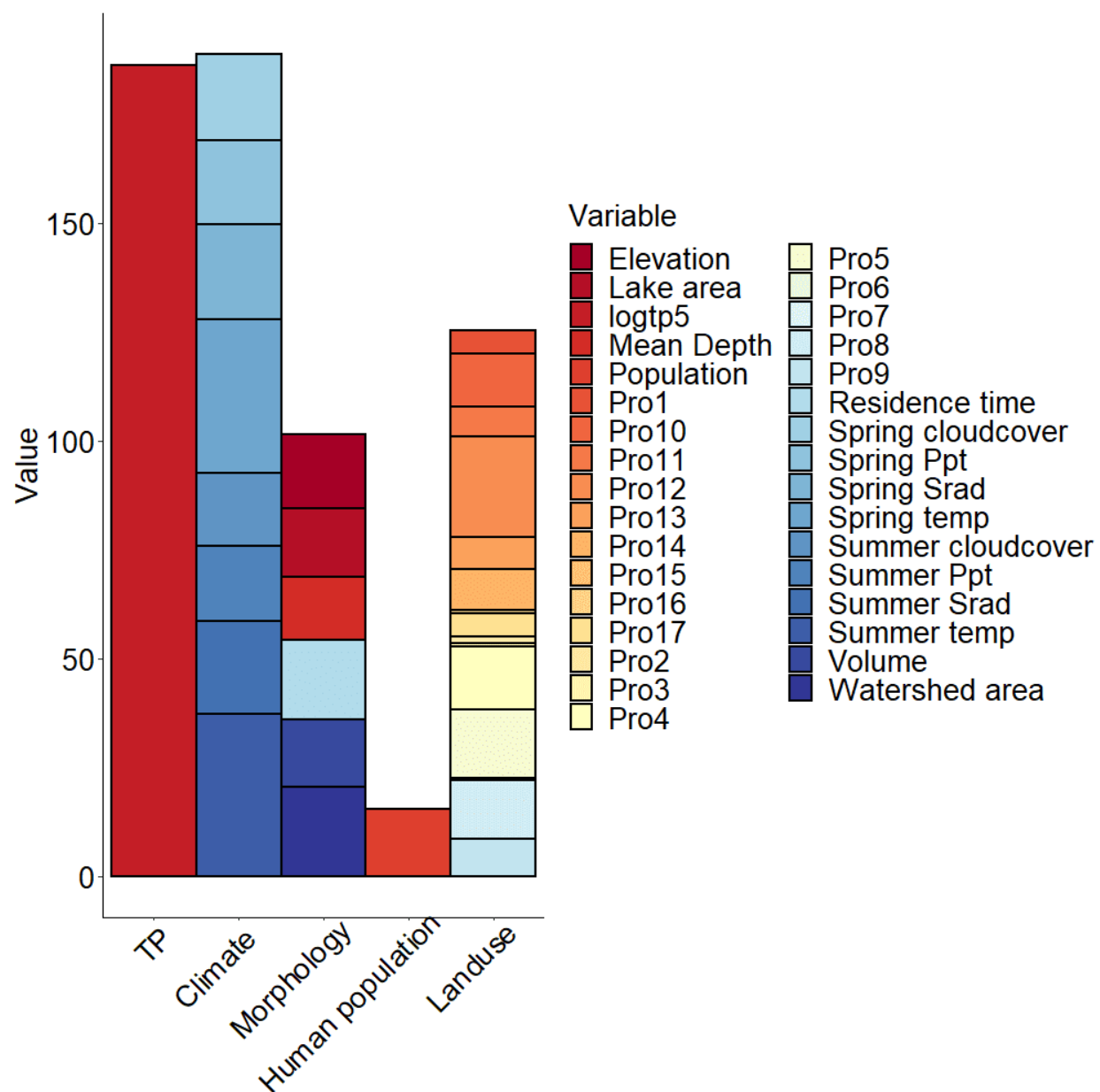


Figure S1. Random forest model (2240 lakes dataset) assessing the key drivers of chlorophyll in lakes across the United states. The drivers assessed were TP, climate, lake morphology, human population, and land use. The random forest model explained 61% variation and the mean of squared residuals was 0.1109.

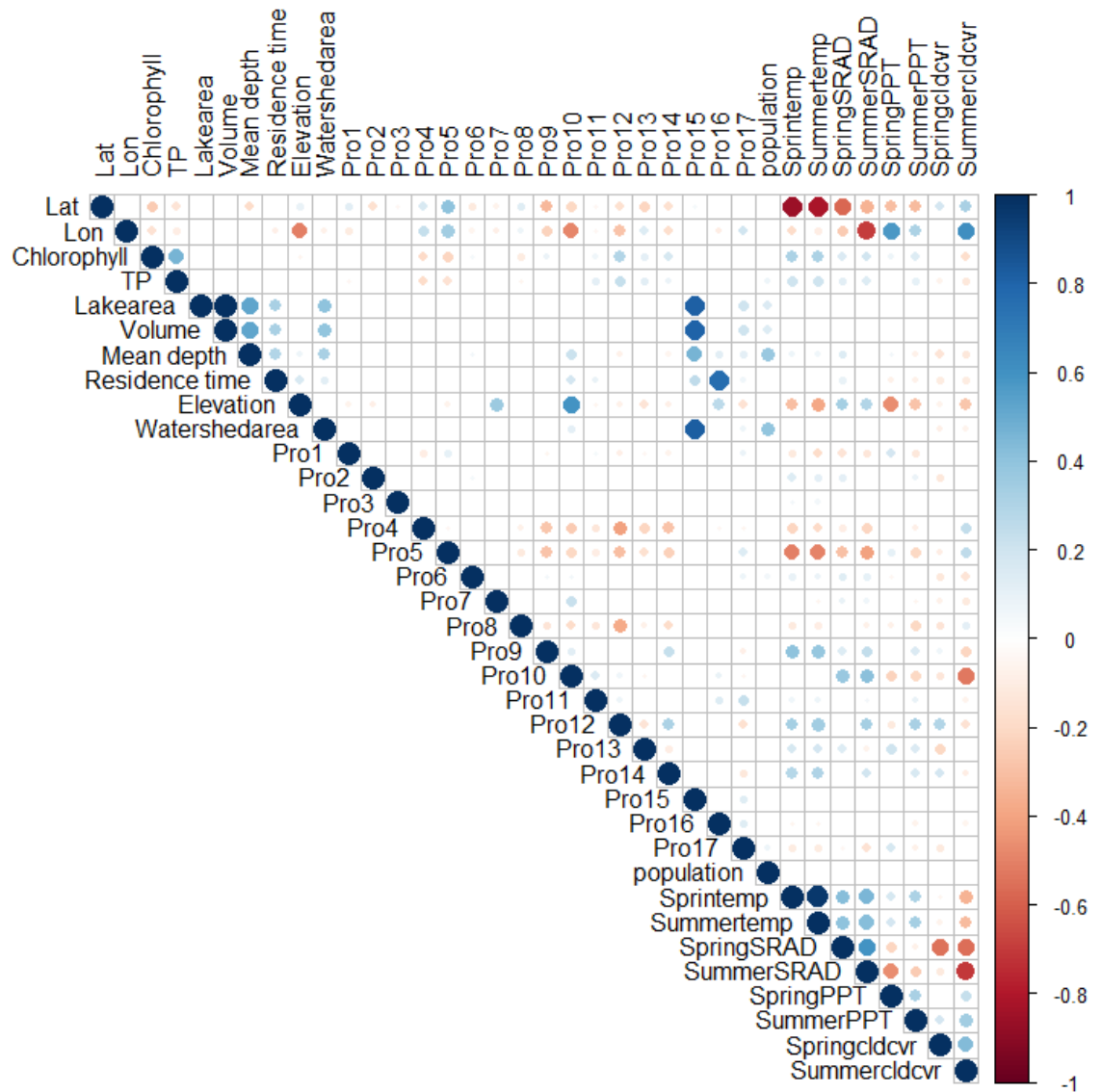


Figure S2. Correlation plot (2240 lakes dataset) highlighting the relationship between chlorophyll (response variable) and total phosphorus (TP), climate, land use, morphology, and human population (predictor variables). Relationships between predictors variables are also shown. Values closer to +1 signify a strong positive relation, whereas -1 suggests a strong negative relationship.

GENERAL CONCLUSION

In Chapter 1, we assessed water quality trends between 3878 lakes globally by examining the relative importance of nutrients inputs, climate conditions, and lake morphology on chlorophyll *a* levels. Our models which incorporated all of the drivers explained 60% of variation in lake chlorophyll *a* levels. Total phosphorus was the most influential driver and explained 42% of the variation. The relationship between TP and chlorophyll has been well established (Schindler, 1977) and has been shown at local and regional scales (Abell et al., 2012), however, our study and Quinlan et al. (in review) shows that this relation also holds true at the global scale. The climactic drivers: air temperature, solar radiation, precipitation, and cloud cover cumulatively explained 38% of the variation. Increases in air temperature and solar radiation were the most influential climactic drivers. The increased air temperatures warmed surface water temperatures, while the increased solar radiation provided optimal light conditions for algal production (O'Reilly et al., 2015; Sharma et al., 2015; Zhong et al., 2016; Collins et al., 2019). Lake geomorphometric drivers were less important compared to nutrient inputs and climate conditions, however, they explained 20% of variation where mean depth and residence time were the most important drivers. Lake depth effects stratification and mixing regimes (Richardson et al., 2017), while residence time encompasses the period in which water is present in a lake prior to evaporation or replacement by new water (Cha et al., 2017). Both mechanisms have been linked to aiding the TP-Chl relationship where shallow lakes have nutrients readily accessible to primary producers (Fergus et al., 2016) and have shown to harbor greater algal biomasses (Huang et al., 2014).

For Chapter 2, we examined the spatial patterns in water quality across the United States in 2240 lakes and identified the relevant spatial signatures of chlorophyll, phosphorus, climate, land use, and morphological variables. Our results suggested that the majority of the lakes in the US are mesotrophic, however over 25% of lakes are eutrophic or hyper-eutrophic. Mesotrophic lakes have moderate productivity owing to their limited nutrient levels, whereas eutrophic and hyper-eutrophic lakes have high levels of biological production and high nutrient levels (Chislock et al., 2013). Water quality in eutrophic/hyper-eutrophics lakes may reach severe degradation where they are no longer able to sustain flora or fauna because of dead zones. These zones have limited oxygen in the water column and there is minimal light penetration beyond surface waters because of large algal mats (Kowalczywska-Madura et al., 2017). Further, we examined the spatial signatures of chlorophyll and its drivers which suggested chlorophyll in locally structured (<500km). This was also the case for TP and morphological variables. Spatial signatures for land use and climate were broader and were found at the watershed (500 – 1500 km) and regional scales (>1500 km) respectively. It was fitting that TP was also similarly locally structured because of the strong TP-Chl relationship and the influence phosphorus has on primary production (Fergus et al. 2016; Quinlan et al., in review). Land use was spatially autocorrelated at the watershed scale – which may be attributed to large scale farming practices or urbanization. Climate was found at the regional scale, but it was interesting to note that the spatial signature for temperature and solar radiation were present at distances >4000 kms. Di Cecco and Gouhier (2019) suggest that the climactic signatures may even be relevant at the global scale and act as drivers or constraints to local drivers (Ward et al. 1998; Miranda et al. 2014; de Oliveira Marcionilio et al. 2015). Further, we found that local drivers (comprised of TP and morphology) were the most influential in structuring lake chlorophyll levels, as they

explained 50% of variation. In addition, TP was the single most important driver of chlorophyll, but cumulatively climatic drivers explained the most variation in lake chlorophyll *a* levels. TP explained 30% of the variation, whereas all significant climate variables, including temperature, precipitation, solar radiation, and cloud cover explained 32%. Land use and morphological drivers were less important compared to TP and climate wherein they explained 20% and 16% respectively. Recent literature supports our findings as they suggest local drivers are more influential than regional drivers (Knowlton & Jones 1997; Floury et al. 2003; Roozen et al. 2003; Gray et al. 2013; Carneiro et al. 2014; Miranda et al. 2014; Marcionilio et al. 2015), however, they are influenced by regional drivers such as land use and climate (Canfield 1983; Miranda et al. 2014; Marcionilio et al. 2015; Quinlan et al., in review).

Both chapters revealed that TP remains the key limiting factor for primary production, and its interactions with other local drivers such as morphometry continues to intensify the impact it has on lake primary production (Sorrano et al., 2010). Geomorphometric and hydrological features were important but had a reduced role relative to climate in terms of explaining overall variation in summer chlorophyll. The relatively similar contributions of climate drivers cumulatively and nutrients suggest that management plans focused primarily on nutrient reductions may also need to consider the implications of ongoing climate change in order to achieve the desired outcomes (Trolle et al., 2011; Paerl et al., 2016). In particular, climatic drivers such as summer temperature and solar radiation act at exceptionally large-scale distances (>4000km). Cecco & Gouhier (2019) suggest that the spatial signature of temperature may even be influential at the global spatial scale and may act as constraints or drivers of the local factors (Ward et al. 1998; Miranda et al. 2014; Marcionilio et al. 2015). As such, we recommend that management efforts and legislation be proposed based on drivers acting at local

spatial scales with consideration of broader regional spatial scales (Soranno et al. 2010; Wagner et al. 2011; Filstrup et al. 2014).

REFERENCES

- Abell, J. M., Özkundakci, D., Hamilton, D. P., & Jones, J. R. (2012). Latitudinal variation in nutrient stoichiometry and chlorophyll-nutrient relationships in lakes: a global study. *Fundamental and Applied Limnology/Archiv für Hydrobiologie*, 181(1), 1-14.
- Canfield Jr, D. E., Bachmann, R. W., & Hoyer, M. V. (1983). Freeze-out of salts in hard-water lakes 1. *Limnology and Oceanography*, 28(5), 970-977.
- Carneiro, F. M., Nabout, J. C., Vieira, L. C., Roland, F., & Bini, L. M. (2014). Determinants of chlorophyll-a concentration in tropical reservoirs. *Hydrobiologia*, 740(1), 89-99.
- Cha, Y., Cho, K. H., Lee, H., Kang, T., & Kim, J. H. (2017). The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. *Water research*, 124, 11-19.
- Chislock, M. F., Doster, E., Zitomer, R. A., & Wilson, A. E. (2013). Eutrophication: causes, consequences, and controls in aquatic ecosystems. *Nature Education Knowledge*, 4(4), 10.
- Collins, S.M., Yuan, S., Tan, P.N., Oliver, S.K., Lapierre, J.F., Cheruvilil, K.S., Fergus, C.E., Skaff, N.K., Stachelek, J., Wagner, T. and Soranno, P.A. (2019). Winter precipitation and summer temperature predict lake water quality at macroscales. *Water Resources Research*, 55(4), 2708-2721.
- Di Cecco, G. J., & Gouhier, T. C. (2018). Increased spatial and temporal autocorrelation of temperature under climate change. *Scientific reports*, 8(1), 1-9.
- de Oliveira Marcionilio, S. M. L., Alves, M. T. R., Borges, P. P., Machado, K. B., Araújo, C. S. T., da Cunha, H. F., & Nabout, J. C. (2015). The state of global scientific literature on chlorophyll-A. *Bioscience Journal*, 31(3).
- Fergus, C. E., Finley, A. O., Soranno, P. A., & Wagner, T. (2016). Spatial variation in nutrient and watercolor effects on lake chlorophyll at macroscales. *Plos one*, 11(10), e0164592.
- Filstrup, C. T., Wagner, T., Soranno, P. A., Stanley, E. H., Stow, C. A., Webster, K. E., & Downing, J. A. (2014). Regional variability among nonlinear chlorophyll—phosphorus relationships in lakes. *Limnology and Oceanography*, 59(5), 1691-1703.
- Gray, B. R., Rogala, J. R., & Houser, J. N. (2013). Treating floodplain lakes of large rivers as study units for variables that vary within lakes; an evaluation using chlorophyll a and inorganic suspended solids data from floodplain lakes of the upper mississippi river. *River Research and Applications*, 29(3), 330-342.
- Huang, J., Xu, Q., Xi, B., Wang, X., Jia, K., Huo, S., Su, J., Zhang, T. and Li, C. (2014). Effects of lake-basin morphological and hydrological characteristics on the eutrophication of shallow lakes in eastern China. *Journal of Great Lakes Research*, 40(3), 666-674.
- Knowlton, M. F., & Jones, J. R. (1997). Trophic status of Missouri River floodplain lakes in relation to basin type and connectivity. *Wetlands*, 17(4), 468-475.
- Kowalczevska-Madura, K., Dondajewska, R., Gołdyn, R., & Podsiadłowski, S. (2017). The influence of restoration measures on phosphorus internal loading from the sediments of a hypereutrophic lake. *Environmental Science and Pollution Research*, 24(16), 14417-14429.
- Miranda, L. E., Andrews, C. S., & Kröger, R. (2014). Connectedness of land use, nutrients, primary production, and fish assemblages in oxbow lakes. *Aquatic sciences*, 76(1), 41-50.
- O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J., Schneider, P., Lenters, J.D., McIntyre, P.B., Kraemer, B.M. and Weyhenmeyer, G.A. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, 42(24), 10-773.

- Paerl, H.W., Gardner, W.S., Havens, K.E., Joyner, A.R., McCarthy, M.J., Newell, S.E., Qin, B. and Scott, J.T. (2016). Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. *Harmful Algae*, 54, 213-222.
- Quinlan, R., A. Filazzola, A., Mahdiyan, O., Shuvo, A., Blagrove, K., Ewins, C., Moslenko, L., Gray, D., O'Reilly, C.M., Sharma, S. (In review). Relationships of total phosphorus and chlorophyll in lake worldwide. *Limnology and Oceanography*
- Richardson, D.C., Melles, S.J., Pilla, R.M., Hetherington, A.L., Knoll, L.B., Williamson, C.E., Kraemer, B.M., Jackson, J.R., Long, E.C., Moore, K. and Rudstam, L.G. (2017). Transparency, geomorphology and mixing regime explain variability in trends in lake temperature and stratification across Northeastern North America (1975–2014). *Water*, 9(6), 442.
- Roozen, F. C., Van Geest, G. J., Ibelings, B. W., Roijackers, R., Scheffer, M., & Buijse, A. D. (2003). Lake age and water level affect the turbidity of floodplain lakes along the lower Rhine. *Freshwater Biology*, 48(3), 519-531.
- Schindler, D. W. (1977). Evolution of phosphorus limitation in lakes. *Science*, 195(4275), 260-262.
- Sharma, S., Gray, D.K., Read, J.S., O'reilly, C.M., Schneider, P., Quadrat, A., Gries, C., Stefanoff, S., Hampton, S.E., Hook, S. and Lenters, J.D. (2015). A global database of lake surface temperatures collected by in situ and satellite methods from 1985–2009. *Scientific data*, 2(1), 1-19.
- Soranno, P. A., Cheruvilil, K. S., Webster, K. E., Bremigan, M. T., Wagner, T., & Stow, C. A. (2010). Using landscape limnology to classify freshwater ecosystems for multi-ecosystem management and conservation. *BioScience*, 60(6), 440-454.
- Trolle, D., Hamilton, D. P., Pilditch, C. A., Duggan, I. C., & Jeppesen, E. (2011). Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management. *Environmental Modelling & Software*, 26(4), 354-370.
- Wagner, T., Soranno, P. A., Webster, K. E., & Cheruvilil, K. S. (2011). Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes. *Freshwater Biology*, 56(9), 1811-1824.
- Ward, J. V., Bretschko, G., Brunke, M., Danielopol, D., Gibert, J., Gonser, T., & Hildrew, A. G. (1998). The boundaries of river systems: the metazoan perspective. *Freshwater Biology*, 40(3), 531-569.
- Zhong, Y., Notaro, M., Vavrus, S. J., & Foster, M. J. (2016). Recent accelerated warming of the Laurentian Great Lakes: Physical drivers. *Limnology and Oceanography*, 61, 1762–1786. doi:10.1002/lno.10331