USING A LANDSCAPE LIMNOLOGY FRAMEWORK TO EXAMINE SPATIAL PATTERNS AND PROCESSES THAT INFLUENCE LAKE NUTRIENTS AND PRODUCTIVITY AT MACROSCALES

By

Carol Emiko Fergus

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Fisheries and Wildlife—Doctor of Philosophy
Ecology, Evolutionary Biology and Behavior—Dual Major

2016

ABSTRACT

USING A LANDSCAPE LIMNOLOGY FRAMEWORK TO EXAMINE SPATIAL PATTERNS AND PROCESSES THAT INFLUENCE LAKE NUTRIENTS AND PRODUCTIVITY AT MACROSCALES

By

Carol Emiko Fergus

Some of the most-pressing and severe environmental perturbations threatening freshwater ecosystems operate at broad spatial scales and are changing at rapid rates such as land use conversion and climate change. However, it is not known how freshwater systems will respond to these broad-scale changes because multi-scaled, geophysical factors promote variation in freshwaters across regional to continental scales and likely influence the effects of these threats on freshwaters. To address these uncertainties requires interdisciplinary perspectives and concepts to study freshwaters within a multi-scaled geophysical context, which the fields of landscape limnology and macrosystems ecology provide. My dissertation research uses these perspectives to examine fundamental questions about spatial variation in and the multi-scaled drivers of lake nutrients and productivity at macroscales.

In chapter 1, I examined spatial variation in the empirical relationships of lake total phosphorus (TP) and water color on chlorophyll *a* in over 800 north temperate lakes using spatially-varying coefficient models to characterize the space-varying and scale-dependent relationships that influence lake primary production. I found spatial autocorrelation in these relationships but that the scale of dependence varied for TP compared to water color.

In chapter 2, I measured lake, wetland, and stream abundance and connectivity at a subcontinental extent in the Midwest and Northeast U.S. to describe macroscale patterns of the freshwater landscape. I found that freshwater abundance and connectivity spatial patterns were

distinct from one another, and these patterns were related to hydrogeomorphic, climate, and land use variables.

In chapter 3, I related multi-scale geospatial features to predict lake total phosphorus, total nitrogen, and water color to evaluate the hypotheses that freshwater systems (i.e., wetlands, streams, and groundwater) and their connectivity influence lake nutrients and carbon concentrations and may promote cross-scale interactions. I found that freshwater systems were related to all three lake response variables, and freshwater systems may promote cross-scale interactions where features at one spatial scale interact with features from another spatial scale to affect lake water chemistry.

The above work allows us to gain a better understanding of the spatial patterns of variation in freshwater systems as well as the underlying variables that may promote this variation across broad spatial extents. Identifying relationships at these macroscales is an important step so that we are better positioned to respond to regional and global change by aligning the spatial scales of how lakes are studied with the scales at which disturbances and management decisions are taking place.

This dissertation is dedicated to my mom, Esther Onaga, my dad, Ted Fergus, and brother, Scott Yukio Fergus, for their support and patience through this endeavor and in all aspects of my life; and to my grandma, Sumie Onaga, who exemplified what it means to work hard and persevere with grace.

ACKNOWLEDGEMENTS

First and foremost I have to thank my major advisor, Dr. Patricia Soranno, who has and continues to be an engaging mentor and role model as a pioneering figure in limnology and macrosystems ecology. Her one-on-one guidance, unique scientific perspective, and leadership of the CSI Limnology research group have shaped not only the chapters presented here but also have profoundly shaped my development as a scientist. In addition, I thank the individual coauthors on the chapters for their contributions and grounding the research in limnological relevancy. I thank my Ph.D. committee, Kendra Cheruvelil, Andy Finley, and Tyler Wagner, for their insightful feedback. I also thank the CSI Limnology research group for providing a supportive and stimulating collaborative environment. And special thanks goes out to Ed Bissell, Nicole Smith, and Scott Stopyak for their hard work in developing the LAGOS database into what it is today.

PREFACE

Each chapter in this dissertation was written as an individual manuscript and is the culmination of separate collaborative efforts with different co-authors. Chapter 1was written with Andrew Finley, Patricia Soranno, and Tyler Wagner and is under review at PlosOne as of July 2016. Chapter 2 was written with Jean-Francois Lapierre, Samantha Oliver, Nicholas Skaff, Kendra Cheruvelil, Caren Scott, Patricia Soranno, and Katherine Webster and will be submitted for review in the fall of 2016. Chapter 3 will be developed into a manuscript in coming months with co-authors that have yet to be determined.

The dissertation chapters presented are my individually-led projects that were embedded within a larger NSF-funded macrosystems ecology research grant, which has yielded multiple side collaborations and coauthored manuscripts during my PhD program. An objective of the overall research project was the creation of LAGOS, a multi-scale geospatial and temporal lake database that spans a subcontinental extent across 17 U.S. states. LAGOS was created by compiling lake water chemistry data from 87 individual lake sampling datasets and geospatial data including climate, atmospheric deposition, land use/land cover, hydrology, geology, and topography measured at multiple spatial scales. In LAGOS, there are lake water chemistry data for close to 10,000 inland lakes and geospatial data for close to 50,000 lakes.

I actively participated in the above project by being a member of the LAGOS-Metadata sub-group and leading the effort to author individual metadata files for each of the 87 limnological datasets that were incorporated into LAGOS. I also was responsible for creating and maintaining documentation for each of the source programs that we received data from, which was incorporated into LAGOS and was a major component of the metadata effort for this large, integrated database.

In addition to the chapters described in this dissertation, and the above work on the project, I have also collaborated with other project members on several additional research articles from 2011 to 2016 that are in different stages of publication. The manuscripts and my specific contributions to the manuscripts are described below.

<u>Published articles</u>

Cross-scale interactions: quantifying multi-scaled cause-effect relationships in macrosystems – *A conceptual and overview article describing the project*

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

- Drafted text for the introduction
- Contributed to analyses, in addition, the results from my 2011 article (Fergus et al.) were highlighted in this article.
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

Building a multi-scaled geospatial temporal ecology database from disparate data sources: fostering open science and data reuse – *A database-methods paper*

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

- Drafted text describing freshwater connectivity metrics and freshwater system geospatial data
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

Prediction of lake depth across a 17-state region in the U.S. – *A research article*

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

- Created maps
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

The statistical power to detect cross-scale interactions at macroscales – *A research article*

Wagner, T., Fergus, C.E., Stow, C.A., Cheruvelil, K.S. & Soranno, P.A. (2016) The statistical power to detect cross-scale interactions at macroscales. *Ecosphere*, **7**, n/a-n/a.

- Drafted text for introduction
- Conducted literature review
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

<u>Publications in review or in preparation</u>

Creating multi-themes ecological regions for macrosystems ecology: Testing a flexible, repeatable, and accessible clustering method – *A research article*

Cheruvelil, K. S., S. Yuan, K. E. Webster, P. N. Tan, J. F. Lapierre, S. M. Collins. C. E. Fergus, C. E. Scott, E. N. Henry, P. A. Soranno, C. T. Filstrup, and T. Wagner. In review. Creating multi-themed ecological regions for macrosystms ecology: Testing a flexible, repeatable, and accessible clustering method.

- Created several figures
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

Climate and land use effects on the spatial patters in CO2 regulation across US lakes – *A research article*

Lapierre, J. F., D. Seekell, C. T. Filstrup, S. M. Collins, C. E. Fergus, P. A. Soranno, K. S. Cheruvelil. *In review*. Climate and land use effects on the spatial patterns in CO_2 regulation across US lakes.

- Conducted literature review
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

LAGOS-NE: a dataset of lake water quality in 17 U.S. states – *A data article*

Soranno, P.A., and many authors. *In preparation*. LAGOS-NE: a dataset of lake water quality in 17 U.S. states.

- Played a major role in creating the database that is described in this data paper
- Contributed during conference calls and meetings to discuss the framing of the manuscript
- Reviewed and edited drafts

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Freshwater ecosystems around the globe face broad-scale disturbances such as land use conversion, the spread of invasive species, and global climate change. Understanding how such threats will affect freshwaters and their ecosystems services will be important to accurately predict their potential effect on society. Traditional single-system studies or multiple within-region studies provide great insight on within-system and catchment mechanistic drivers of freshwater ecosystem integrity. However, such studies cannot capture among-region variation nor complex multi-scaled interactions that may govern freshwater ecosystem variation. It is recognized that biological, chemical, physical, and human components interacting across multiple spatial and temporal scales can result in emergent or unexpected system properties at macroscales (Michener *et al.*, 2001; Heffernan *et al.*, 2014). And this ecological complexity makes it challenging to extrapolate relationships to new areas and understand how lakes may respond to global change. Therefore, there is a need for spatially-explicit, system approaches to study freshwater bodies within the context of the surrounding complex and multi-scaled landscape.

The emerging sub-disciplines of landscape limnology and macrosystems ecology offer theoretical frameworks to build conceptual understanding of the multi-scale drivers that influence freshwater composition and ecological functions. Here I apply landscape limnology and macrosystems ecology approaches to address pressing questions in freshwater ecology on lake nutrients and primary production by studying three important components of variation in freshwaters at macroscales: (1) spatial variation in the nutrient and carbon drivers of lake productivity; (2) the underlying spatial patterns in the abundance and connectivity of the entire freshwater landscape that includes lakes, streams, and wetlands; and (3) the role of cross-scaled interactions in explaining lake nutrient variation.

Collectively, the following chapters use a dual landscape limnology/macrosystems ecology approach to further our understanding of freshwater systems at macroscales. This approach examines patterns and processes at multiple scales to provide insight on the factors that influence lake ecosystem composition and function. The results can inform both fundamental and applied scientific questions to better understand and manage freshwater systems.

Chapter 1 – Spatial variation in lake ecosystem processes

The nutrient-water color paradigm is a framework to better characterize lake trophic status by relating lake primary productivity to both nutrients and water color, the colored component of dissolved organic carbon, rather than considering nutrients alone (Williamson *et al.*, 1999; Webster *et al.*, 2008). The positive relationship between total phosphorus and lake chlorophyll a (an indicator of lake primary productivity) has been documented extensively in the literature; however, the magnitude of this relationship varies from study to study. Additionally, water color effects on lake primary production are highly variable and can be in contrasting directions. Thus it is not known how well the nutrient-water color paradigm applies to broader populations of lakes. It is expected that nutrient and water color relationships are spatially structured because the hypothesized drivers of variation (e.g., catchment geomorphology, landscape sources of nutrients and carbon, and community composition) vary across the landscape. However, few studies have examined spatial heterogeneity of total phosphorus and water color effects together on lake chlorophyll a, and this variation may provide insight on the underlying geospatial variables that promote differences in these relationships. Chapter 1 examines spatial variation in total phosphorus and water color relationships in over 800 north temperate lakes using spatially-varying coefficient models to characterize the space-varying and scale-dependent relationships that influence lake primary production.

Chapter 2 – Patterns of freshwater abundance and connectivity at macroscales

From a landscape limnology perspective, lakes and other freshwater ecosystems are influenced by multi-scaled terrestrial, aquatic, human, and climatic features in the surrounding landscape (Soranno *et al.*, 2009, 2010). Reciprocally, it is becoming widely recognized that freshwater systems, including lakes, wetlands, and streams, have large impacts on regional and global processes such as carbon and nutrient budgets (Cole *et al.*, 2007; Tranvik *et al.*, 2009). However, freshwater systems cannot be integrated into macroscale processes without first an understanding of the distribution and abundance of lakes, wetlands, and streams at macroscales. Regional to global freshwater abundance estimates have been conducted for lakes (Downing *et al.*, 2006), wetlands (Lehner & Döll, 2004), and streams (Butman & Raymond, 2011) individually. But, abundance measures are coarse representations of freshwater systems and their potential contributions to macroscales processes. Abundance measures do not capture spatial configuration among systems that are known to be important for ecological functions. Freshwater connectivity, defined as the surface hydrologic connections that link lakes, wetlands, and streams, mediates the transport of water, materials, nutrients, organisms, and energy across the landscape, and thus is likely to influence the role of freshwater systems in macroscale processes. However, we have a poor understanding of the patterns of freshwater connectivity at broad spatial scales. In Chapter 2, I measured lake, wetland, and stream abundance and connectivity at a subcontinental extent in the Midwest and Northeast U.S. to study macroscale patterns of the freshwater landscape. I present a robust approach to quantify freshwater abundance and connectivity to integrate freshwater ecosystems into macroscale studies.

Chapter 3 – Cross-scale interactions and freshwater connectivity

Cross-scale interactions (CSIs), defined as "processes at one spatial or temporal scale interacting with processes at another scale to result in nonlinear dynamics" (Peters *et al.*, 2007) are one of the defining macrosystems interactions that promote ecosystem complexity (Heffernan *et al.*, 2014; Soranno *et al.*, 2014). It is expected that CSIs may play an influential role in lake ecosystem behavior, but there are few examples of CSIs in freshwater systems (but see Fergus *et al.*, 2011; Wagner *et al.*, 2011; Filstrup *et al.*, 2014). Part of the lack of studies on CSIs may be due to the absence of a strong theoretical framework that identifies characteristics and mechanisms of cross-scale interactions. Connectivity has been proposed as a key attribute promoting interactions across spatial and temporal scales (Peters *et al.*, 2008). These connections are the transport mechanisms that may moderate the influence of regional components on lake ecosystems and the potential for cross-scale interactions to be in operation. While connectivity associations with cross-scale interactions have been examined in terrestrial systems (e.g., Falk *et al.*, 2007; Young *et al.*, 2007), these associations have not been explicitly studied in aquatic systems. For lakes, freshwater connections may be strongly associated with CSIs because they are physical conduits that distribute water, materials, nutrients, organisms, and energy among geographically separated landscape elements. These connections may amplify or attenuate CSIs depending on the type of ecological process. But the link between connectivity and CSIs is still poorly defined. In Chapter 3, I examine freshwater connectivity and multi-scale geographic features to predict lake total phosphorus, total nitrogen, and water color. These analyses will provide insight on the potential role of freshwater connectivity in promoting cross-scale interactions in lake ecosystems.

Chapter 1: Spatial variation in nutrient and water color effects on lake primary production at macroscales

Abstract

The nutrient-water color paradigm is a framework to characterize lake trophic status by relating lake primary productivity to both nutrients and water color, the colored component of dissolved organic carbon. Total phosphorus (TP), a limiting nutrient, and water color, a strong light attenuator, influence lake chlorophyll *a* concentrations (CHL). But these relationships have been shown in previous studies to be highly variable which may be related to differences in lake and catchment geomorphology, the forms of nutrients and carbon entering the system, and lake community composition. Because many of these variables vary across space it is unclear how well the nutrient-water color paradigm applies to lakes distributed across diverse landscape settings. Although it is expected that nutrient and water color relationships are spatially structured, few studies have examined spatial heterogeneity in the effects of TP and water color together on lake CHL. In this study, we examined spatial variation in TP and water color relationships with CHL in over 800 north temperate lakes using spatially-varying coefficient models (SVC), a robust statistical method that applies a Bayesian framework to explore space-varying and scale-dependent relationships. We found that allowing for TP and water color relationships to vary over space improved the model fit and predictive performance over models that did not vary over space. The magnitudes of TP and water color effects on CHL varied across lakes and the spatial scales of variation of these two drivers were different for our study lakes. For example, the variation in TP–CHL relationships occur at intermediate distances (~20 km)

compared to variation in water color–CHL relationships that occur at regional distances (~200 km). These results demonstrate that the effects of nutrient and water color on lake primary production are influenced by spatially structured features that may operate at different spatial scales and that quantifying spatial structure in TP and water color effects on lake chlorophyll furthers our understanding of the variability in these relationships at macroscales.

Introduction

A longstanding goal in limnology and lake management is to develop empirical models to predict lake primary production from nutrient concentrations. However, these models can exhibit a great deal of variation in predictive performance across studies. The nutrient-water color paradigm has been proposed to help account for these differences by recognizing that lake trophic condition is characterized by measures of *both* nutrients and water color (the colored component of dissolved organic carbon) in contrast to examining nutrients alone (Nürnberg & Shaw, 1998; Williamson *et al.*, 1999; Webster *et al.*, 2008). Lake primary production measures, such as chlorophyll *a* (CHL), have been shown to be strongly related to both phosphorus, a limiting nutrient in temperate North American lakes, and water color, a strong light attenuator, but in potentially contrasting directions (Mazumder, 1994; Carpenter *et al.*, 1998b; Nürnberg & Shaw, 1998; Phillips *et al.*, 2008). However, because few studies have examined the effects of phosphorus and water color on lake CHL together (Nürnberg & Shaw, 1998; Webster *et al.*, 2008), it is unclear how well the nutrient-water color paradigm applies to lakes distributed across diverse landscape settings. Understanding the relative strength of these drivers of lake productivity will be especially important to predict how lakes may respond to ongoing and future global changes that are altering nutrient and carbon inputs to freshwater systems (Heathwaite *et al.*, 1996; Bennett *et al.*, 2001; Driscoll *et al.*, 2003).

While the empirical relationship between CHL and lake total phosphorus (TP) is well established, no consensus has been reached on the additional role of water color on lake primary production. Water color can influence lake primary production in complex and confounding ways (Carpenter *et al.*, 1998b; Nürnberg & Shaw, 1998; Karlsson *et al.*, 2009; Seekell *et al.*, 2015a). The strong light attenuating effects of water color can inhibit photosynthesis and reduce phytoplankton abundance (Carpenter *et al.*, 1998b; Thrane *et al.*, 2014). In contrast, water color has also been positively associated with primary production by directly supplying nutrients to aquatic systems. Humic substances can form complexes with nutrients and thus be sources of inorganic nutrients to lakes (Jones, 1992; Klug, 2002; Kissman *et al.*, 2013). Water color can also stimulate primary production by indirectly promoting processes that release nutrients. Spectral properties of water color can influence the mixing depth in small lakes (Fee *et al.*, 1996) and subsequently promote biogeochemical conditions to release nutrients from the sediment, which can stimulate primary production (Brothers *et al.*, 2014). It is difficult to study water color effects across large populations of lakes because the contrasting effects of these different mechanisms could cancel out the overall effect on primary production, and it is likely that these mechanisms operate at different spatial and temporal scales.

There are several lines of evidence that show that TP and water color relationships with CHL are spatially structured and influenced by landscape features. First, land use and land cover are major sources of both phosphorus and carbon to lakes and these landscape features vary across space. For example, agriculture land use and wetland cover near lakes are recognized sources of phosphorus and dissolved organic carbon, respectively, to lakes (Xenopoulos *et al.*, 2003; Taranu & Gregory-Eaves, 2008). Second, comparative studies at broad spatial extents demonstrate that the effect of TP on CHL varies across ecoregion units (Wagner *et al.*, 2011; Filstrup *et al.*, 2014). Similarly, dissolved organic carbon relationships with primary production exhibit strong among-region differences (Wagner *et al.*, 2011; Filstrup *et al.*, 2014; Seekell *et al.*, 2015a). And finally, other spatially structured features, such as topography and geology, may influence in-lake processing of nutrients and carbon (Canham *et al.*, 2004; Phillips *et al.*, 2008; Webster *et al.*, 2008) and consequently affect primary production, leading to further spatial structuring of the relationships between driver and response variables. It should be noted that lake community composition (herbivore assemblages and macrophyte coverage) has been related to differences in TP and CHL relationships (Canfield Jr. *et al.*, 1984; Mazumder, 1994; Carpenter *et al.*, 1998b), and these biological attributes are likely to exhibit spatial patterns but potentially with a high level of variation from lake to lake.

Although it is expected that there is spatial structuring of TP and water color relationships with lake CHL, it can be difficult to explicitly examine and account for these spatial dependencies. Previous studies accounted for spatial variation in driver and response relationships by using discrete spatial units, such as ecoregions, to partition the landscape into ecologically similar patches and capture variation in lake response variables (Phillips *et al.*, 2008; Wagner *et al.*, 2011; Filstrup *et al.*, 2014). While these discrete spatial units improve model accuracy, they may not optimally delineate the landscape to capture spatial variation in TP–CHL and water color–CHL relationships. In other words, variation in these relationships may occur at finer spatial extents than the ecoregion boundaries chosen for the above analyses. In addition, there may be scale differences among TP and water color effects such that the spatial drivers of TP–CHL relationships may operate at different spatial scales than the spatial drivers of water color–CHL relationships; and confining variation to fixed ecoregion boundaries may not be the best scale for either TP nor water color.

In this paper we explore the nutrient-water color paradigm for over 800 lakes across diverse landscape settings to examine spatial heterogeneity in CHL relationships with TP and water color at broad spatial extents. We ask the following questions, 1) Do TP and water color relationships with CHL vary among lakes at sub-continental scales? And 2) If so, at what spatial scale do these relationships vary (Gelfand *et al.*, 2004)? To answer these questions, we fit spatially-varying coefficient (SVC) models using a Bayesian framework to lakes located in the Midwest and Northeast regions of the U.S. The SVC model allows for regression model intercept and slope parameters (i.e., coefficients) to vary over continuous space rather than among discrete regional units. Specifically, each regression coefficient is modeled using Gaussian process, with mean, variance, and distance correlation decay parameters estimated using a valid probability model. With this modeling approach we can quantify the spatial range where spatial dependency in parameter values diminish and identify spatial scales that capture variation in TP and water color relationships separately. We included covariates in the models that have been shown to be related to lake CHL (e.g., lake depth) and assessed their influence on lake CHL. Finally, we explored whether the lake-specific spatially-varying coefficients were related to hypothesized lake and catchment characteristics using correlation analyses. In general, quantifying spatial variation in TP and water color relationships with lake CHL at macroscales should improve model inference and provide insight on the relative strength of nutrient and water color drivers on lake primary production.

Methods

Lake and landscape datasets

Data used in the analyses come from the LAGOS database (Lake multi-scaled geospatial and temporal database, (Soranno *et al.*, 2015)). LAGOS is a multi-thematic lake database that integrates lake water chemistry data ($LAGOS_{LIMNO}$) and geospatial data ($LAGOS_{GEO}$) across the U.S. Midwest and Northeast regions. We accessed $LAGOS_{LIMNO}$ version 1.040.0 and $LAGOS_{GEO}$ version 1.02 for this study.

For our analyses we used a subset of lakes from the LAGOS dataset with water chemistry and lake geomorphology data related to our research questions. Our dataset included lakes (greater than 4 ha and less than 10,000 ha in surface area) that had summer (15 June – 15 September) epilimnetic CHL, TP, and water color observations measured at the same sampling event. We omitted lakes missing maximum depth data from our analysis because lake depth is recognized to affect nutrient processing and primary production and it was an important variable to include in the models. In total, the data included 838 lakes (7395 observations) within Wisconsin (WI), Michigan (MI), New York (NY), and Maine (ME) (Fig 1) and captured a wide range of lake and catchment characteristics (Table 1). Data and metadata for this study are available at the Long-Term Ecological Research Network Data Portal doi: 10.6073/pasta/0ebd2e4c0705706b77b359955bff44e1 (Fergus *et al.*, 2016).

Lake CHL, TP, and water color data were collected by state agencies from 1986 – 2013 following standard field collection and laboratory methods. The majority of lakes in the dataset (>70%) have a single water chemistry observation over time. There are several lakes (N = 228) with multiple observations over time, and these are mainly located in New York (Appendix Fig. A1). Lakes with multiple observations had, on average, about 30 observations, with most of the observations occurring across years (i.e., not within the same season each year). We checked for temporal autocorrelation in water chemistry measurements for individual lakes by examining residual plots over time and did not find evidence for either among-year trends or within-year (seasonal) trends that would need to be accounted for in the model design. Thus, we kept multiple water chemistry observations per lake over time in the dataset to increase the number of observations used to fit the models.

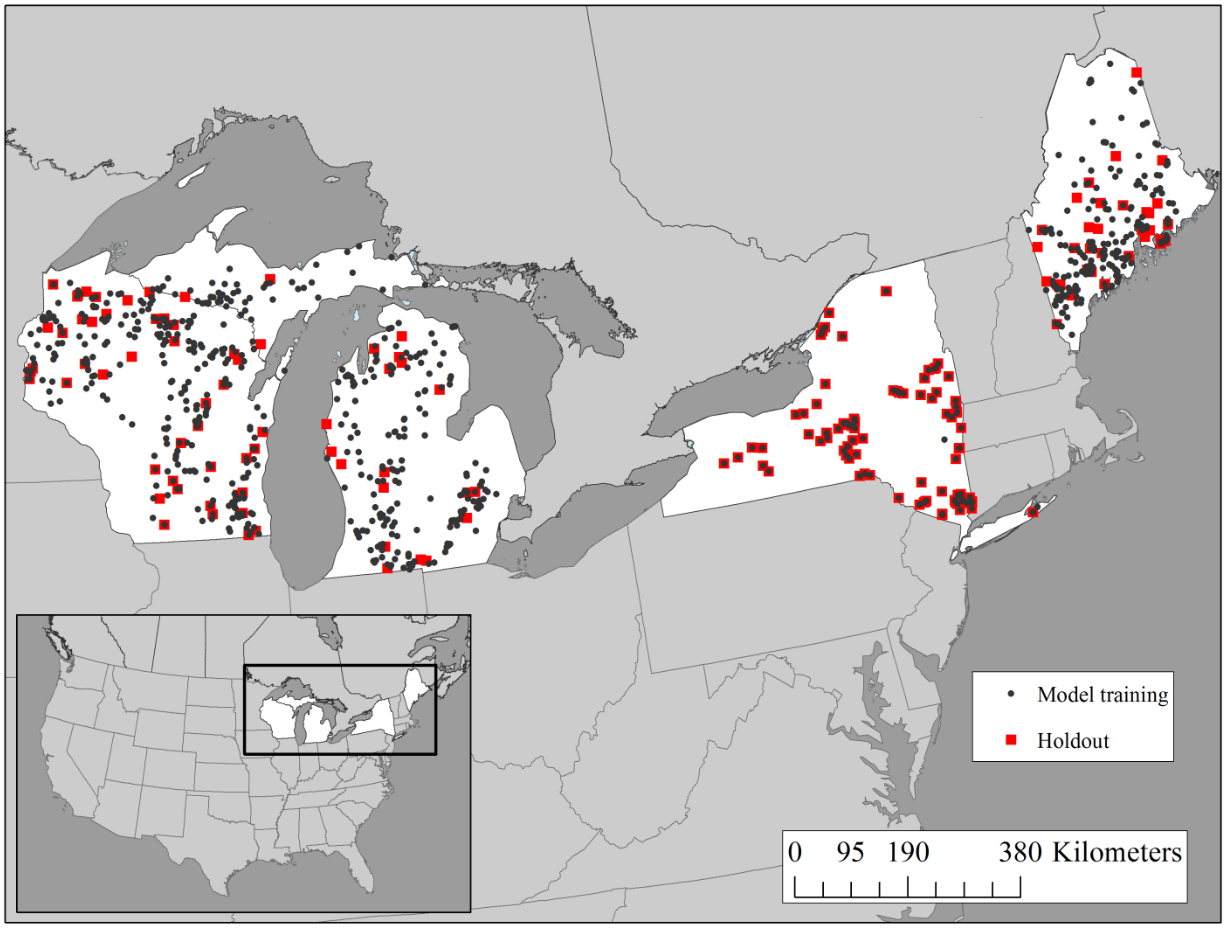

Figure 1. **Study extent map.** Lake locations in the analysis (N = 838 lakes) including model training observations (n = 6656) and locations of holdout observations for model predictive performance (n = 739).

Table 1. Summary statistics of the full lake dataset. The mean, median, range, and standard deviation of lake water chemistry, lake geomorphology, and landscape variables for the full dataset (n = 7395 observations, N = 838 unique lakes). Prop. = proportion in the lake catchment. CA:LK = catchment to lake area ratio.

Variable	Mean	Median	Range	Standard deviation
Chlorophyll *a* (μg/L)	10.72	4.47	0.01 – 363.00	19.23
TP (μg/L)	21.97	14.00	0.90 – 494.00	27.07
Water color (PCU)	20.30	14.00	1.00 – 194.00	21.05
Max. depth (m)	11.54	9.20	1.52 – 58.50	8.23
Lake area (ha)	230.00	55.49	4.28 – 7043.36	578.38
Catchment area (ha)	4976.00	654.70	3.90 – 436923.90	19394.01
CA:LK	26.06	10.06	0.27 – 7444.23	113.32
Prop. Agriculture	0.17	0.10	0 – 0.84	0.18
Prop. Urban	0.10	0.05	0 – 0.96	0.15
Prop. Wetland	0.10	0.06	0 – 0.81	0.11
Prop. Forest	0.10	0.06	0 – 0.63	0.10

We related lake CHL and spatially-varying coefficients to lake hydrogeomorphology and catchment variables in $\mathrm{LAGOS}_{\mathrm{GEO}}$. In $\mathrm{LAGOS}_{\mathrm{GEO}}$, lakes were assigned a hydrologic connectivity type based on the presence or absence of surface stream connections represented in the National Hydrography Dataset (NHD) (see (Soranno *et al.*, 2015) for methods used to identify lake hydrologic type). Lakes were identified as either *isolated* (i.e., no inflowing streams) or *drainage* (i.e., inflowing streams). In the dataset there were N = 606 *drainage* lakes (with 5896 observations) and N = 232 *isolated* lakes (with 1499 observations). Mean and standard deviation values of lake and catchment variables by lake type are available in Appendix Table A1. Catchment boundaries were delineated for each lake in the study extent using automated Geographic Information Systems (GIS) methods (see LAGOS GIS Toolbox, (Soranno *et al.*, 2015)). Land use and land cover class proportions within the lake catchments were quantified from the 2006 National Land Cover Database because the majority of water chemistry data were collected around this year.

Analysis

Model framework overview

We applied SVC models within a Bayesian inferential framework to examine variation in TP and water color relationships with CHL over space. SVC models are suited to our research questions because they allow for the explicit examination of both space-varying and scale-dependent relationships between nutrient and color drivers and lake chlorophyll. SVC models allow for selected model regression coefficients to vary by point locations and produce smoothly varying coefficient surfaces that are modeled as realizations from spatial processes (Gelfand *et al.*, 2003) and because of this the models do not assume that coefficients are stationary (i.e., constant) over space – allowing for inference about location specific impact of drivers on the response. In contrast to multi-level mixed effects models that use discrete areal units to model spatial dependency, SVC models allow for greater flexibility and relieve the constraint of modeling variation among potentially arbitrarily-specified areal units that may not optimally capture the scale of spatial variation across the different covariates. The Bayesian framework produces posterior probability distributions that allow for full uncertainty quantification in parameter estimates and subsequent predictions at unobserved locations within the domain.

Model description

SVC model structure took on the following form. We model log CHL $y_t(\boldsymbol{s})$ at lake location $\boldsymbol{s}$ and sample time t as

$$y_t(\boldsymbol{s}) = \tilde{\boldsymbol{x}}_t(\boldsymbol{s})^{\mathrm{T}}\,\tilde{\boldsymbol{\beta}}(\boldsymbol{s}) + \boldsymbol{x}_t(\boldsymbol{s})^{\mathrm{T}}\,\boldsymbol{\beta} + e_t(\boldsymbol{s})$$

Where $\tilde{\boldsymbol{x}}_t(\boldsymbol{s})$ is an intercept with lake and time specific measurements of log TP and log water color, i.e., $\tilde{\boldsymbol{x}}_t(\boldsymbol{s}) = (1, TP_t(\boldsymbol{s}), color_t(\boldsymbol{s}))^{\mathrm{T}}$, and $\tilde{\boldsymbol{\beta}}(\boldsymbol{s})$ is the associated vector of spatially-varying regression coefficients. Additional covariates with spatially invariant regression coefficients are specified in $\boldsymbol{x}_t(\boldsymbol{s})$ and $\boldsymbol{\beta}$, respectively. Model residuals are assumed to follow a zero-centered normal distribution that is independent across measurement location and time, i.e., $e_t(\boldsymbol{s}) \sim N(0, \tau^2)$ where , τ^2 is the residual variance parameter that captures measurement error. We assume $\tilde{\boldsymbol{\beta}}(\boldsymbol{s})$ follows a multivariate Gaussian process, i.e., MVGP $\left(\tilde{\boldsymbol{\beta}}_{mu}, \sum(\boldsymbol{\theta})\right)$ where $\tilde{\boldsymbol{\beta}}_{mu}$ is the mean regression coefficients over the domain and $\sum(\boldsymbol{\theta})$ is the covariance matrix with $\boldsymbol{\theta}$ including an intercept, TP, and water color specific spatial correlation decay parameters (φ) and cross-covariance parameters. The MVGP is constructed using a Linear Model of Coregionalization (see, e.g., (Gelfand *et al.*, 2004)).

We quantified the distance at which the spatial dependence in model coefficient values becomes negligible by calculating the effective spatial range. The effective spatial range is based on the spatial correlation decay parameters (φ). We define the effective spatial range as the distance at which the spatial correlation drops to 0.05 between observations (Finley, 2011). The effective spatial range provides an estimate of the spatial scale that captures variation in lake TP and water color effects on CHL.

We evaluated four hypothesized candidate models to examine the potential spatially structured effects of TP and water color on lake CHL. The first candidate model was a non-spatial linear regression relating TP and water color to CHL that estimated global model coefficients that were fixed across the dataset. The second model ($\mathrm{SVC}_{\mathrm{TP,COLOR}}$) allowed the intercept, TP, and water color regression coefficients to vary spatially. The third model

($SVC_{LANDSCAPE}$) had the same spatially varying coefficients (i.e., intercept, TP, and water color) and also included hypothesized lake (maximum depth, catchment to lake area ratio – CA:LK) and landscape (proportion agriculture and wetland area in the catchment) space invariant covariates. These covariates were included in the models because they have been related to lake primary production and water chemistry concentrations in the literature (Rasmussen *et al.*, 1989a; Kortelainen, 1993; Webster *et al.*, 2008) and they did not exhibit strong collinearity with one another. The fourth model (SVC_{FULL}) built upon the $SVC_{LANDSCAPE}$ model by including a dummy variable to identify the lake connectivity type (*isolated* vs. *drainage*). We $\log_{10}$ transformed CHL, TP, water color, maximum depth, and CA:LK to reduce skewness of the data.

Model evaluation and predictive performance

The candidate non-spatial and SVC models were evaluated two ways: (1) model fit to the data and (2) predictive performance using out-of-sample cross-validation. Prior to model building, 90% of the observations (n = 6656) in the dataset were selected at random and used to estimate candidate models' parameters, and the remaining 10% of observations (n = 739) were withheld to evaluate model predictive performance. To evaluate the fit of the candidate models to the observed data, we used the deviance information criterion (DIC), an information criterion that can be used to compare models that apply a Bayesian framework (Spiegelhalter *et al.*, 2002). DIC is calculated as the sum of the Bayesian deviance value (D) and estimated effective number of parameters in the model (pD), where lower DIC values indicate better model fit. For the out-of-sample cross-validation, the parameter posterior samples for the model-fitting dataset were used to generate posterior predictive samples for the holdout observations (see (Banerjee *et al.*, 2014)). Then, using the holdout observations and model posterior predictive distribution samples, predictive performance was summarized using 1) root mean-square prediction error

(RMSPE) between observed values and means of the predictive distributions; 2) mean continuous rank probability score (CRPS), which is a strictly proper scoring rule that quantifies the fit of the entire predictive distribution (i.e., for a normal distribution, the mean and the variance) to the data (Gneiting & Raftery, 2007); 3) percent of observations covered by their corresponding predictive distribution 95% credible interval (PCI) and mean width of the predictive distributions' 95% credible interval (PIW). Lower values of RMSPE and CRPS indicate better predictive performance. Similarly, we would favor models that provide narrow posterior predictive interval widths (PIW) while delivering appropriate posterior coverage rates, i.e., PCI at ~95%.

Finally, we explored whether spatial variation in TP and water color relationships were related to underlying lake and catchment characteristics using Pearson correlation analyses. We related the mean posterior coefficient values (i.e., lake specific intercept, TP, and water color slopes) estimated from the $SVC_{TP,COLOR}$ model (the model that did not include any of the fixed-effect lake and landscape covariates) to hypothesized lake and catchment variables. Mean differences in spatially-varying coefficient values among lake connectivity types were assessed using Welch two-sample t-tests. Correlation and t-test analyses were performed using base packages in R statistical platform (R Core Team 2015).

Results

The SVC models that included spatially-varying intercept, TP, and water color coefficients were better models in terms of fit and predictive performance compared to the non-spatial model. Among the SVC models, the top ranked model was SVC_{FULL} that included spatially-varying intercept, TP, and water color coefficients in addition to spatially-invariant lake and landscape coefficients (Table 2). The DIC and D values were the lowest for SVC_{FULL} compared to the other candidate models indicating a better model fit to the observed data despite being penalized for including additional parameters in the model. In terms of model predictive performance, the SVC models performed similarly well and provided improved RMSPE and CRPS over the non-spatial regression and acceptable coverage rates with a narrower PCI (Table 2).

Lake and landscape covariates modeled as spatially-invariant (i.e., fixed across lake location)

The top-ranked model based on DIC values included spatially-varying coefficients and lake and landscape variables that were modeled as stationary, or to have fixed effects across locations. The lake and landscape covariates that were important in the model (based on coefficient 95% credible intervals not overlapping zero) had expected relationships with CHL (Table 2). Maximum lake depth was negatively associated with CHL – such that deeper lakes tended to have lower CHL concentrations in comparison to more shallow lakes. Proportion of agricultural activity in the catchment and lake connectivity type were positively associated with CHL. Lakes with high agricultural land use in their catchments had higher CHL concentrations compared to lakes with less agricultural activity. Including a dummy variable to indicate lake connectivity type indicated that drainage lakes had higher CHL compared to isolated lakes. Catchment area to lake area ratio (CA:LK) and proportion of wetland cover in the catchment showed no discernable relationship with CHL. It should be noted that the lack of relationships in this model does not mean that these covariates are not important, but rather that the effect of these covariates may vary across locations in our study extent and may be difficult to detect.

Table 2. Summary of TP and water color ~ CHL candidate models including posterior estimated coefficients, model fit criteria, and model predictive performance measures. Model coefficient posterior means are presented with 95% credible intervals. The residual variance parameter (τ^2) quantifies measurement error. The effective spatial range values (km) are calculated for the spatially-varying coefficients based on spatial decay parameters φ_1, φ_2, φ_3. Models are ranked based on deviance information criteria (DIC) scores where lower values indicate a better model fit. The effective number of parameters (pD) are taken into account in the DIC scores (based on Bayesian deviance value D) to penalize more complex models. Model predictive performance is summarized using root mean-square predictive error (RMSPE), mean continuous rank probability score (CRPS), percent of observations covered by their corresponding predictive distribution 95% credible interval (PCI), and mean width of the predictive distributions' 95% credible interval (PIW). Smaller RMSPE and CRPS values indicate better predictive performance, larger PCI values indicate increased model accuracy, and smaller PIW indicate increased precision.

	Non-spatial	**$SVC_{TP,COLOR}$**	**$SVC_{LANDSCAPE}$**	**SVC_{FULL}**
Intercept ($\tilde{\beta}_0$)	-1.12	-0.43	-0.34	-0.36
	(-1.20, -1.04)	(-0.57, -0.29)	(-0.61, -0.08)	(-0.61, -0.10)
TP ($\tilde{\beta}_{TP}$)	1.06	0.73	0.71	0.698
	(1.03, 1.09)	(0.67, 0.79)	(0.64, 0.77)	(0.63, 0.77)
Color ($\tilde{\beta}_{Color}$)	-0.06	-0.002	-0.01	-0.02
	(-0.09, -0.04)	(-0.086, 0.094)	(-0.10, 0.07)	(-0.10, 0.08)
Z_{MAX} (β_{Zmax})			-0.08	-0.13
			(-0.16, -0.01)	(-0.19, -0.04)
CA:LK (β_{CALK})			0.05	0.02
			(0.01, 0.09)	(-0.03, 0.07)
AGR (β_{AGR})			0.52	0.45
			(0.18, 0.83)	(0.13, 0.77)
WET (β_{WET})			0.29	0.16
			(-0.25, 0.87)	(-0.38, 0.73)
Drain. (β_{Drain})				0.21
				(0.08, 0.34)
τ^2	0.82	0.63	0.63	0.63
	(0.79, 0.85)	(0.60, 0.65)	(0.61, 0.65)	(0.61, 0.65)
Eff. Range $_{Intercept}$		21.78	14.21	32.56
		(19.33, 26.37)	(12.81, 15.71)	(19.90, 119.15)
Eff. Range $_{TP}$		19.98	33.75	26.32
		(17.91, 23.58)	(27.47, 39.66)	(16.78, 99.41)
Eff. Range $_{Color}$		302.07	442.43	216.05
		(199.53, 443.56)	(311.81, 537.24)	(138.98, 276.62)
ΔDIC	1457.09	16.65	27.75	0
pD	3.94	324.54	310.41	322.08
D	10887.70	8348.76	8388.56	8329.50
RMSPE	0.88	0.77	0.77	0.77
CRPS	0.48	0.42	0.42	0.42
95% PCI	94.74	90.68	90.23	90.25
95% PIW	3.52	2.51	2.48	2.45

Spatially-varying model coefficients

Spatially-varying intercept

Spatially-varying intercepts and TP and water color effects (i.e., slopes) on CHL improved model fit compared to the non-spatial, regression model. This indicated that lake chlorophyll concentrations in addition to the effects of phosphorus and water color on lake chlorophyll exhibited spatial autocorrelation across a diverse set of north temperate, inland lakes. Allowing for the model intercept values to vary across lake locations captured spatial autocorrelation in lake CHL that was not accounted for by the mean function and subsequently helped meet the model assumptions that residuals are independent and identically distributed. Lake and landscape covariates added to models $SVC_{LANDSCAPE}$ and SVC_{FULL} smoothed remaining spatial variation in lake CHL as seen in the spatially-varying intercept surface maps (Fig. 2).

Spatially-varying TP and water color effects on CHL (i.e., spatially-varying slopes)

The SVC_{FULL} model estimated the mean effects of TP and water color on CHL; however, it is more informative to examine the spatially-varying coefficients across locations to better understand the distribution of these coefficient values over the study extent. The spatial processes that captured variation in model coefficients were different for TP compared to water color suggesting that there may be different underlying factors that influence TP and water color effects on CHL. Lake TP was positively related to CHL for all lakes, but the magnitudes of these effects were different across locations (Fig 3). The posterior mean log CHL–log TP coefficient for the study lakes was 0.73 (±0.84) and ranged from 0.27 to 1.38. Translating these values into effects on CHL, on average 1 µg/L increase in TP was related to 5 µg/L increase in lake CHL.

However, TP~ CHL effects were variable across individual lakes with a 1 µg/L increase in TP resulting in increased CHL ranging from 2 – 24 µg/L.

In contrast, water color effects on CHL varied over space but were not important for the majority of lakes in the study extent (Fig 4). Where water color effects were significant (interpreted as the SVC 95% credible intervals not overlapping zero), some lakes had positive water color relationships with CHL (mean SVC_{Color} 0.30 ± 0.02; N = 4 lakes) and other lakes had negative water color relationships with CHL (mean SVC_{Color} -0.26 ± 0.07; N = 16 lakes).

The scales at which coefficients were spatially structured were different for TP–CHL effects compared to water color–CHL effects based on the effective spatial range values. In the SVC_{FULL} model, the effective spatial range for TP–CHL was 26 km (Table 2) indicating that lakes within 26 km of one another had more similar TP relationships with CHL compared to lakes that were further away. In contrast, the effective spatial range for water color–CHL was 216 km (Table 2). Because the majority of lakes did not have significant water color relationships with CHL, this large effective spatial range indicates that there are broad spatial areas where lakes had weak to non-existent water color relationships with CHL (Fig 4b).

Once we established that spatially-varying coefficients improved the model fit, we examined whether hypothesized lake and landscape variables were related to spatial variation in these effects using the spatially-varying coefficients from the $SVC_{TP,COLOR}$ model. Lake-specific TP–CHL coefficients were not strongly correlated with maximum lake depth or catchment characteristics ($r < 0.5$; Table 3). There was a statistically significant difference in the mean TP–CHL coefficient values between drainage and isolated lakes (two sample t-test: $t = 2.55$; $df = 441.31$; p-value < 0.05); but the difference in mean TP–CHL effects between lake types was small and does not appear to be ecologically meaningful (Appendix Fig. A2).

Similarly, lake-specific water color–CHL coefficients were not strongly correlated with any of the hypothesized lake and catchment characteristics ($r < 0.5$; Table 3). There was a statistically significant difference in the mean water color–CHL coefficient values between drainage and isolated lakes (two sample t-test: $t = 2.40$; $df = 428.40$; p-value < 0.05); but again the difference was small and may not be ecologically important (Appendix Fig. A2).

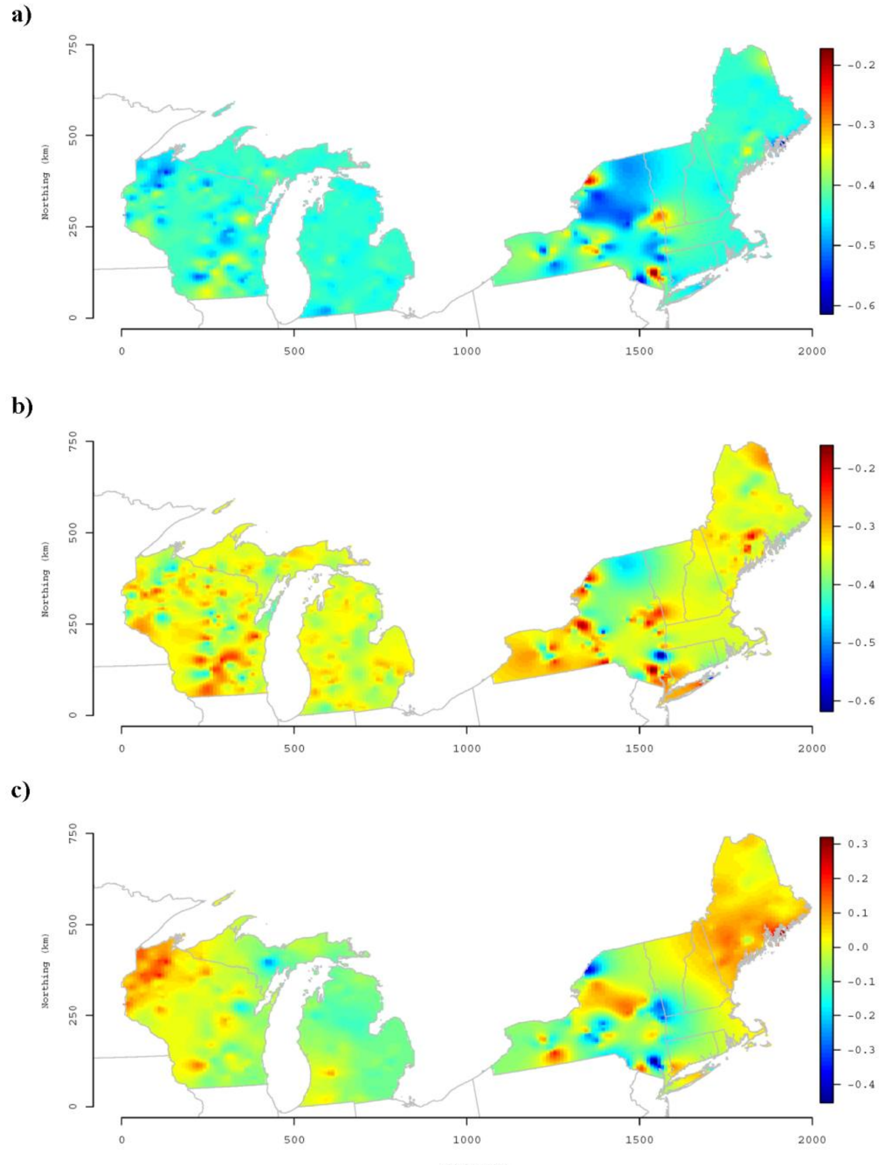

Figure 2. Spatially-varying intercept surface maps for a) $SVC_{TP,COLOR}$, b) $SVC_{LANDSCAPE}$, and c) SVC_{FULL} models. Interpolated surface maps were derived from the posterior mean of the spatially-varying intercept values estimated by lake location in the model building dataset (N = 779) and displayed as blue to red color gradients representing low to high intercept values.

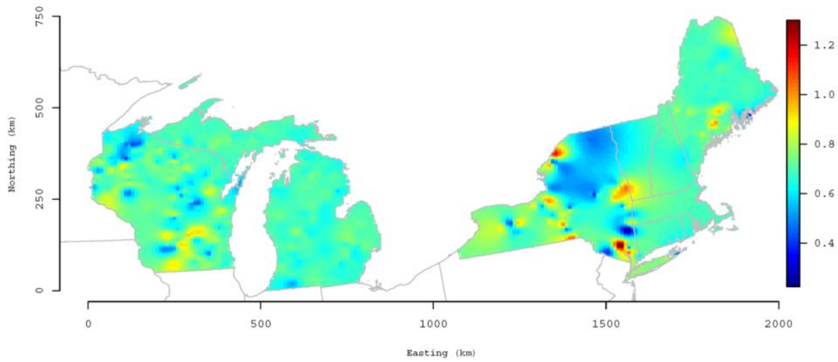

Figure 3. Spatially-varying TP–CHL coefficients maps derived from the SVC_{FULL} model. a) Surface map of spatially-varying TP–CHL relationships created by interpolation of the posterior mean values that were estimated by lake location in the model building dataset (N = 779). Blue to red color gradient represents low to high TP–CHL coefficient values.

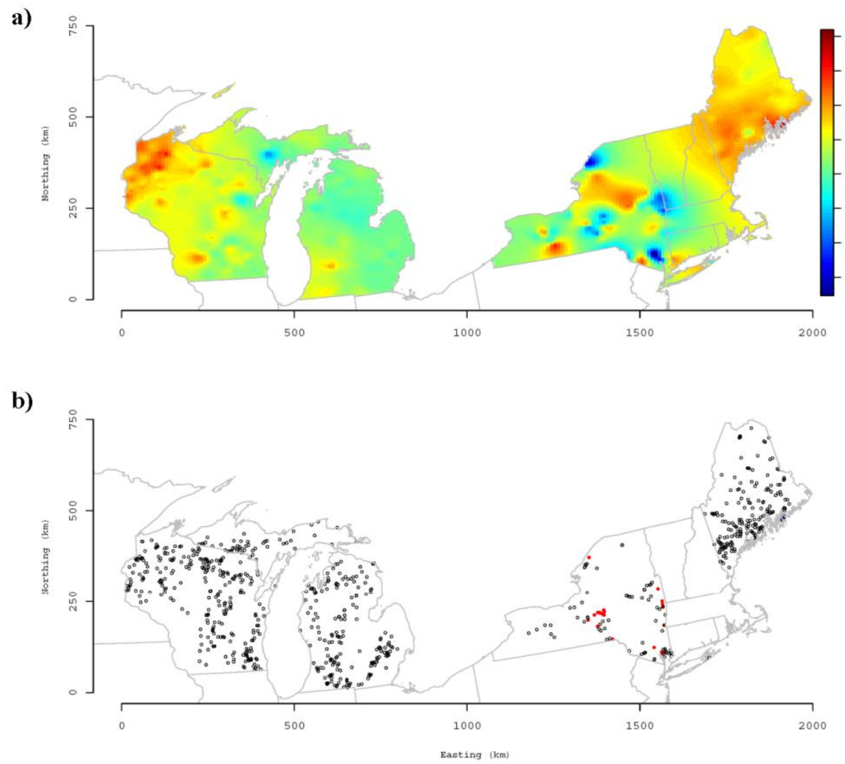

Figure 4. Spatially-varying water color–CHL coefficients maps derived from the SVC$_{FULL}$ model. a) Surface map of spatially-varying water color–CHL relationships created by interpolation of the posterior mean values that were estimated by lake location in the model building dataset (N = 779). Blue to red color gradient represents low to high water color–CHL coefficient values. b) Map of lake point locations symbolized by water color–CHL relationships: positive (blue), negative (red), not significant (black outlined dot). Significant relationships were determined based on 95% credible intervals not overlapping zero.

Table 3. Correlation coefficient values for lake-specific spatially-varying coefficients and hypothesized lake and catchment variables. Spatially-varying Intercept ($SVC_{INTERCEPT}$), TP (SVC_{TP}), and water color (SVC_{COLOR}) coefficients were estimated for 779 lakes from the $SVC_{TP,COLOR}$ model. Significant correlation coefficients ($\alpha < 0.05$) are in bold.

	$SVC_{INTERCEPT}$	SVC_{TP}	SVC_{COLOR}
log_{10}-Secchi	**-0.30**	**-0.38**	**0.20**
log_{10}-Zmax	**-0.11**	**-0.14**	**0.15**
log_{10}-CALK	**0.12**	**0.12**	**-0.08**
AG	**0.13**	**0.15**	**-0.08**
WET	0.01	0.01	**-0.13**

Discussion

Our results demonstrate that lake water chemistry relationships with primary production measures (i.e., TP–CHL and water color–CHL) exhibit potentially important lake-to-lake differences that are spatially structured at broad extents. Modeling spatial autocorrelation in TP and water color relationships improved inference (based on DIC scores) and prediction (based on RMSPE) over the model that ignored spatial dependency and provided insight on the spatial characteristics of these relationships. The spatial scales that structure TP–CHL relationships were different from the spatial scales that structure water color–CHL relationships. Specifically, variation in TP effects on CHL was structured at a more local scale (~20 km), which means that lakes within a 20 km radius, have similar TP–CHL relationships. In contrast, variation in water color effects was structured at a more regional scale (~200 km). This is the first study to our knowledge that explicitly examined spatial variation over continuous space of the well-recognized lake TP–CHL relationship and the highly variable water color–CHL relationship. The results further our understanding of the multi-scaled structure of nutrient and water color relationships that control lake primary production (i.e., the nutrient-color paradigm) and offer insight in identifying appropriate spatial scales for limnological research and water resource management.

Spatially-varying TP–CHL relationships

We found that TP–CHL relationships exhibited a great deal of spatial variation in our study extent. The lake-specific log TP–CHL slopes (0.27 – 1.38) are within the range of values reported in the literature (Phillips *et al.*, 2008). Several studies have tried to improve TP–CHL predictions by evaluating sources of variation in these relationships, but few studies have examined spatial variation in TP–CHL relationships (Phillips *et al.*, 2008; Wagner *et al.*, 2011; Filstrup *et al.*, 2014). Wagner et al. (2011) found regional differences in TP–CHL relationships within ecological drainage units (EDU) that range in area from 1,000 km^2 to 10,000 km^2 (Higgins *et al.*, 2005) such that lakes within regions had more similar TP effects compared to lakes from other regions. In another study, no regional differences in TP–CHL relationships were detected among coarsely delineated regions for European lakes (Phillips *et al.*, 2008). This lack of any regional relationship may have been because the regions spanned multiple countries in the European Union and may have captured a great deal of within-region heterogeneity. Our results on TP–CHL relationships by lake suggest that variation in these relationships occur at intermediate spatial scales between lake catchment and commonly used ecoregion extents (e.g., EDU). Regional delineations that are ~400 km^2 in area may more optimally capture variation in TP–CHL relationships over space compared to larger regional extents. Future studies may want to consider using intermediate-sized spatial extents to capture variation in TP–CHL relationships compared to using broad ecoregional extents alone.

We hypothesized that the spatial variation in TP–CHL relationships estimated from the

SVC models would be related to lake geomorphic and catchment characteristics (Soranno *et al.*, 2010). However, we did not find evidence for any strong associations. The lack of any associations may be due to scale differences in the response and the predictor variables selected (Wu, 2004). The spatial scales that landscape variables were quantified (i.e., catchment scale) were not aligned with the spatial scales of variation in TP–CHL relationships and thus exhibited weak correlations. Alternatively, differences in TP effects may be influenced by complex, cross-scale interactions where features at one spatial scale may interact with features at another scale [34–36]. In fact, there is evidence for cross-scale interactions being associated with differences in TP–CHL relationships in other studies. Regional percentage of pasture land was associated with among-region differences in TP relationships with CHL, illustrating an example of features at one spatial scale (i.e., region) interacting with processes at another spatial scale (i.e., lake) (Wagner *et al.*, 2011). Similarly Filstrup et al. (2014) found that the percentage of pasture and wetlands within the region were related to TP–CHL effects modeled as nonlinear relationships (Filstrup *et al.*, 2014). These findings suggest that there may be broader landscape features beyond the lake catchment that structure differences in TP effects on CHL and a multi-scaled perspective is warranted.

At the opposite end of the spatial continuum, the variation observed in TP – CHL relationships across the study lakes may be related to a number of in-lake characteristics such as differences in morphology, water chemistry, and zooplankton and macrophyte community composition (Canfield Jr. *et al.*, 1984; Mazumder, 1994; Phillips *et al.*, 2008; Yuan & Pollard, 2014). We did not find support that maximum lake depth was associated with spatial differences in TP–CHL relationships. However, spatial differences in these relationships may be related to unmeasured water chemistry variables and lake community composition characteristics that are linked to landscape sources and spatial dispersion factors (Kling *et al.*, 2000; Beisner *et al.*, 2006).

Total nitrogen and total phosphorus ratios (TN:TP) and alkalinity have been related to variation in TP–CHL relationships and are tightly linked to land use activity and geological composition in the landscape. Lakes with very low TN:TP ratios have been related to reduced TP–CHL relationships due to N-limitation (Prairie *et al.*, 1989; Downing & McCauley, 1992). The forms of agricultural activity can influence TN:TP ratios. For example, row-crop activity is associated with high TN:TP ratios and pasture is associated with low TN:TP ratios (Arbuckle & Downing, 2001). In our study, total nitrogen data were not available for most of our study lakes, but we distinguished between agriculture NLCD classes (cultivated land vs. pasture) and used these classes as indicators of nutrient ratios exported to lakes. We did not see a strong correlation between agriculture type and lake-specific TP effects (cultivated land $r = 0.11$ and pasture $r = 0.17$). However, it should be noted that there are some uncertainties in agricultural land use class specification (Wickham *et al.*, 2013). The weak relationship may also be attributed to the temporal land cover period not being well-aligned with the water chemistry data. Alkalinity of lakes is associated with decreased chlorophyll yield per unit of phosphorus due to phosphorus precipitating out of solution (Håkanson *et al.*, 2005). However other studies show no strong association among geological indicators of alkalinity and variation in TP–CHL (Wagner *et al.*, 2011). We lacked data on alkalinity for our study lakes to properly explore this relationship but it is worth investigating in future studies.

Lake community composition has also been related to differences in TP–CHL relationships. Large zooplankton herbivore communities have been associated with lower CHL yields per unit TP across different lake trophic classes (Mazumder, 1994). And increased macrophyte coverage was associated with lower lake chlorophyll production (Canfield Jr. *et al.*, 1984). Macrophyte and zooplankton community composition in lakes may be structured by spatial factors that influence dispersal such as hydrologic connectivity (Beisner *et al.*, 2006) and may be related to spatial variation in TP – CHL relationships. However we did not have lake community composition data for our study lakes and it may be that these spatial factors would operate at finer spatial scales than the intermediate spatial scales observed.

Spatially-varying water color–CHL relationships

Although the majority of lakes did not exhibit significant water color effects on CHL, there were some lakes that had positive water color effects and some lakes that had negative effects. The lake with a maximum positive water color effect on CHL resulted in a 2.14 μg/L increase in CHL per unit increase in water color. The lake with the greatest negative water color effect resulted in a 2.67 μg/L decrease in CHL per unit increase in water color. It was expected that water color effects on lake chlorophyll would not be significant in the global model because these contrasting positive and negative relationships would cancel one another out. However, we expected to find more individual lakes with significant color effects than what was observed. The results suggest that water color effects on lake CHL may be less important compared to TP effects for north temperate lakes in areas with mixed land use/cover. Regional patterns of lake organic carbon coupled with nonlinearities in DOC relationships with primary production may be why we did not detect a strong water color relationship in our study extent. Lake DOC and water color concentrations are shown to exhibit regional patterns that are related to underlying landscape and climatic features (Fergus *et al.*, 2011; Seekell *et al.*, 2014; Lapierre *et al.*, 2015). In addition, in northern boreal and arctic lakes, dissolved organic carbon is shown to have a nonlinear relationship with lake primary productivity such that at low concentrations DOC is positively associated with primary production (acting as a nutrient source by carrying P) and at high concentrations it is negatively associated with primary production by inhibiting light availability (Seekell *et al.*, 2015a,b). Our study lakes did not capture a wide range of water color and the distribution was skewed towards low colored lakes (<20 PCU). Thus these low concentrations might result in weak positive effects on CHL, but these effects would be washed out in disturbed landscapes where there are more prolific landscape nutrient sources (e.g., agriculture).

There were too few lakes with significant water color relationships to draw definitive conclusions on what promotes differences in water color effects on CHL, but here we describe the general characteristics of these lakes to identify potential lake and catchment variables to explore in future studies. Lakes that exhibited significant positive water color–CHL relationships were deep, oligotrophic lakes with catchments dominated by forest cover and minimal human disturbances (Appendix Table A2). These lakes had moderate wetland cover in their catchments, but of this total wetland area the majority of patches were connected to streams in the catchment suggesting a potential mechanisms of carbon transport to the lake. Connected wetland patches may be important sources of colored dissolved organic carbon to these lakes (Laudon *et al.*, 2011). In contrast, lakes that exhibited significant negative water color–CHL relationships tended to be less deep, mesotrophic lakes with moderate levels of agricultural land use in the catchment. These patterns might suggest that land use disturbance influences the relationship between water color and lake chlorophyll, but further studies are needed.

Spatial variation in lake CHL

Allowing lake CHL (i.e., model intercept) to spatially vary by lake improved the model fit to the observed data. Even after accounting for TP and water color effects, lake CHL exhibited spatial heterogeneity that was structured at intermediate scales with an effective range of ~30 km (Table 2). This indicated that lakes that are within 30 km of one another have more similar CHL compared to lakes that are located further away and that there may be underlying spatially-structured variables that promote these patterns of CHL. We found that lake and catchment predictor variables included in the top-ranked model improved model fit of observed CHL and accounted for some of the spatial variation in lake CHL (i.e., model intercepts) indicating that these predictor variables themselves exhibit spatially heterogeneity that promoted the spatial patterns of CHL observed in the study lakes. These variables followed expected relationships with CHL with maximum lake depth having a negative effect on CHL and proportion agricultural land use and lake connectivity type having a positive effect on CHL.

Lake depth is recognized as an important lake geomorphological characteristic that influences in-lake physical, chemical, and biological processes such as mixing regime, water residence time, and nutrient dynamics (Kalff, 2002). Deep lakes tend to have lower total phosphorus and water color concentrations compared to shallow lakes (Rasmussen *et al.*, 1989b; Taranu & Gregory-Eaves, 2008; Webster *et al.*, 2008), which can decrease primary production. While lake depth does not appear to exhibit strong spatial autocorrelation at broad spatial extents (Sobek *et al.*, 2011), topographic features in the surrounding landscape are related to lake depth (Heathcote *et al.*, 2015) suggesting that it may exhibit some spatial structure, which may be related to the spatial variation observed in CHL for our study lakes.

The proportion of agricultural land use in the catchment was positively associated with CHL. Agricultural land use is a recognized nonpoint nutrient pollution source to lakes that can subsequently influence primary production in lakes (Carpenter *et al.*, 1998a). Agricultural activities in the landscape exhibit non-random spatial patterns related to underlying topographical and geological features that constrain locations of land use change (Pan *et al.*, 1999). Additionally, nutrient loadings to the catchment from different agricultural practices have been shown to exhibit distinctive spatial heterogeneity (Puckett, 1995). Together these spatial characteristics of agricultural land use may account for the observed lake CHL spatial patterns.

Lake connectivity type was related to CHL such that drainage lakes had higher concentrations of CHL compared to isolated lakes. Lake connectivity groups had distinguishing lake and catchment characteristics that may promote differences in primary production. Drainage lakes had larger catchments compared to isolated lakes (median = 1355.91 ha vs. median = 200.71 ha) and a greater amount of agriculture in the catchment compared to isolated lakes (median = 6% vs. median = 3%), which may promote differences in CHL concentrations among lake types.

We did not find significant relationships for neither CA:LK nor the proportion of wetlands in the catchment. CA:LK may be a poor indicator of important lake-landscape processes that influence primary production. It was not surprising that there was no relationship between wetland cover and lake CHL. Wetlands have complex relationships with nutrient dynamics and primary production in lakes such that they can have confounding effects among lakes across broad spatial extents (Fergus *et al.*, 2011). Modeling wetland effects as a global mean effect may not be appropriate for these macroscale analyses due to regional differences in the effects of wetlands on lakes. However, our main focus was on evaluating differences in TP and water color effects over space so we did not allow for wetland effects to vary by lake.

It should be noted that there was remaining spatial variation in lake CHL that was not accounted for by the predictors included in the model. However, the effective spatial range for the model intercept (~ 30 km) can assist in identifying potential landscape variables that are structured at similar scales and may account for CHL variation.

In conclusion, quantifying spatial structure in TP and water color effects on chlorophyll helps to expand our understanding of the variability in these relationships, which define the nutrient-color paradigm, over broad spatial extents and diverse lake types. As more focus turns toward adopting macroscale frameworks to address global change at broad scales, there is a need for innovative analytical approaches that can allow for spatial dependencies in such data. SVC models are one approach to improve model prediction and quantify spatial scales of variation in complex ecological relationships.

APPENDIX

Chapter 1: Appendix Tables & Figures

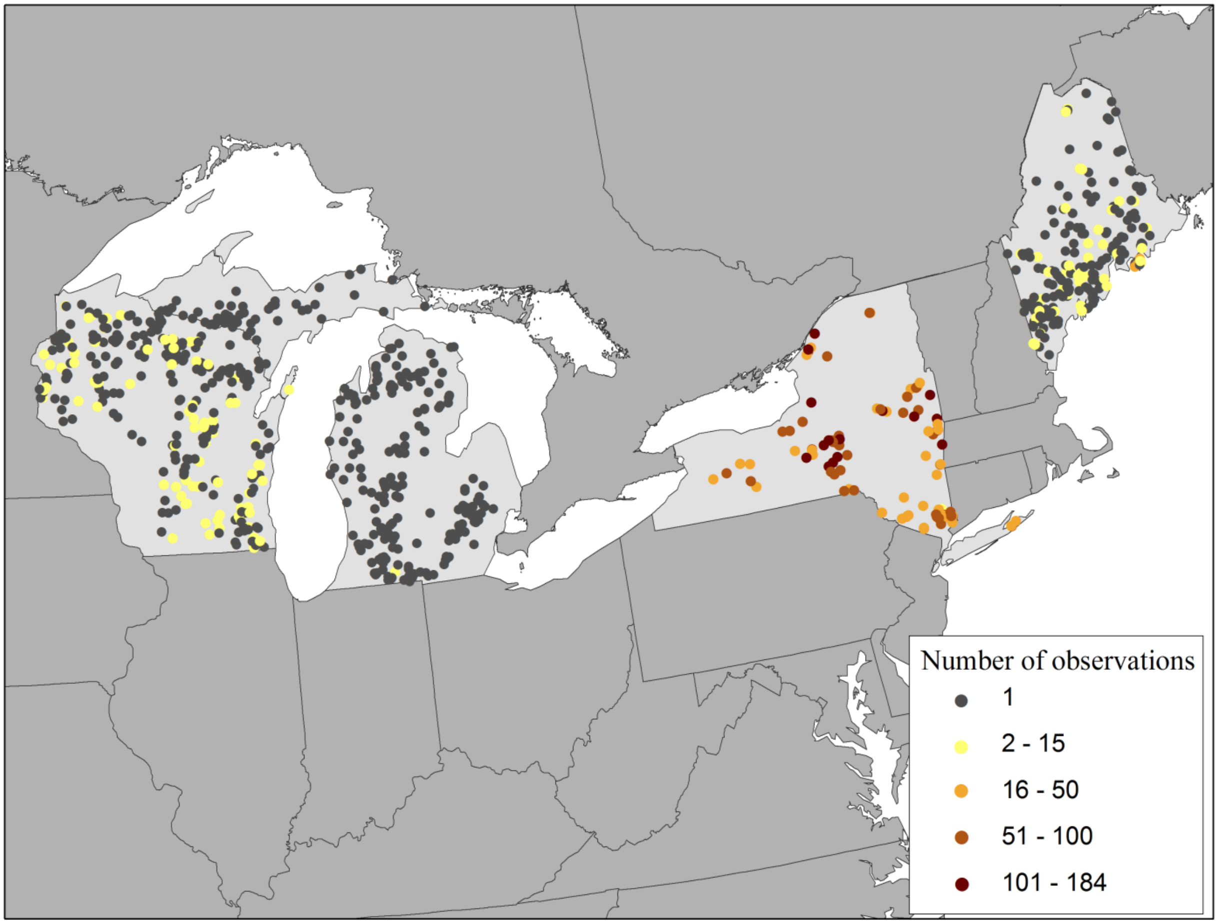

Figure A1. Map of study extent and distribution of number of water chemistry observations by lake. Lakes are symbolized by number of water chemistry observations with gray representing single observations and a yellow to red color gradient representing multiple water chemistry observations by lake.

Table A1. Mean and standard deviation values by lake connectivity type. Mean and standard deviation (sd) values for lake water chemistry and lake and catchment characteristics were quantified by lake connectivity type in the full dataset (N = 838 lakes).

Variable	Isolated	Drainage
Chlorophyll a (μg/L)	12.29 (21.47)	10.32 (18.60)
TP (μg/L)	23.83 (27.20)	21.49 (27.03)
Water color (PCU)	20.40 (20.54)	20.27 (21.17)
Max. depth (m)	9.43 (8.00)	12.07 (8.20)
Lake area (ha)	43.93 (89.04)	277.31 (637.60)
Catchment area (ha)	256.91 (360.63)	6175.76 (21555.39)
CA:LK	8.51 (9.29)	30.53 (126.44)
Prop. Agriculture	0.05 (0.13)	0.20 (0.19)
Prop. Urban	0.14 (0.19)	0.09 (0.14)
Prop. Wetland	0.08 (0.10)	0.11(0.11)
Prop. Forest	0.65 (0.26)	0.57 (0.22)

Table A2. Characteristics of lakes with positive and negative water color – CHL relationships. Summary statistics on lake water chemistry variables and hypothesized lake and landscape covariates for lakes with significant positive water color–CHL relationships (N = 4 lakes) and significant negative relationships (N = 16 lakes). Prop. = proportion in the lake catchment.

Variable	Mean	Median	Range	Standard deviation
Positive water color				
Chlorophyll a (μg/L)	1.37	1.29	1.02 – 1.90	0.37
TP (μg/L)	4.63	4.71	2.10 – 7.00	2.04
Water color (PCU)	5.62	6.00	2.00 – 8.50	3.14
Max. depth (m)	27.81	26.82	11.89 – 45.72	14.90
Lake area (ha)	93.30	85.58	13.36 – 188.66	72.58
Catchment area (ha)	446.30	486.8	179.20 – 632.40	197.66
CA:LK	7.14	6.11	2.91 – 13.41	4.46
Prop. Agriculture	0	0	0 – 0	0
Prop. Urban	0.08	0.08	0 – 0.11	0.03
Prop. Wetland	0.01	0.02	0 – 0.02	0.01
Prop. Forest	0.71	0.72	0.58 – 0.80	0.01
Negative water color				
Chlorophyll a (μg/L)	10.09	5.61	1.69 – 44.37	11.04
TP (μg/L)	18.65	13.12	6.62 – 42.75	12.28
Water color (PCU)	14.59	9.62	5.00 – 50.37	12.28
Max. depth (m)	10.56	9.40	4.00 – 23.30	5.49
Lake area (ha)	267.58	39.42	5.12 – 3179.20	780.71
Catchment area (ha)	8196.50	1081.20	61.30 – 104456.20	25765.36
CA:LK	54.82	19.09	3.93 – 434.84	112.08
Prop. Agriculture	0.21	0.21	0 – 0.45	0.15
Prop. Urban	0.08	0.05	0 – 0.46	0.11
Prop. Wetland	0.08	0.06	0 – 0.43	0.10
Prop. Forest	0.55	0.52	0.32 – 0.91	0.19

LITERATURE CITED

LITERATURE CITED

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**, 970–975.

Banerjee, S., Carlin, B.P. & Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*, CRC Press.

Beisner, B.E., Peres-Neto, P.R., Lindström, E.S., Barnett, A. & Longhi, M.L. (2006) The Role of Environmental and Spatial Processes in Structuring Lake Communities from Bacteria to Fish. *Ecology*, **87**, 2985–2991.

Bennett, E.M., Carpenter, S.R. & Caraco, N.F. (2001) Human Impact on Erodable Phosphorus and Eutrophication: A Global Perspective Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. *BioScience*, **51**, 227–234.

Brothers, S., Köhler, J., Attermeyer, K., Grossart, H.P., Mehner, T., Meyer, N., Scharnweber, K. & Hilt, S. (2014) A feedback loop links brownification and anoxia in a temperate, shallow lake. *Limnology and Oceanography*, **59**, 1388–1398.

Canfield Jr., D.E., Shireman, J.V., Colle, D.E., Haller, W.T., Watkins II, C.E. & Maceina, M.J. (1984) Prediction of Chlorophyll a Concentrations in Florida Lakes: Importance of Aquatic Macrophytes. *Canadian Journal of Fisheries and Aquatic Sciences*, **41**, 497–501.

Canham, C.D., Pace, M.L., Papaik, M.J., Primack, A.G.B., Roy, K.M., Maranger, R.J., Curran, R.P. & Spada, D.M. (2004) A spatially explicit watershed-scale analysis of dissolved organic carbon in adirondack lakes. *Ecological Applications*, **14**, 839–854.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N. & Smith, V.H. (1998a) Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, **8**, 559–568.

Carpenter, S.R., Cole, J.J., Kitchell, J.F. & Pace, M.L. (1998b) Impact of dissolved organic carbon, phosphorus, and grazing on phytoplankton biomass and production in experimental lakes. *Limnology and oceanography*, **43**, 73–80.

Downing, J.A. & McCauley, E. (1992) The nitrogen : phosphorus relationship in lakes. *Limnology and Oceanography*, **37**, 936–945.

Driscoll, C.T., Driscoll, K.M., Roy, K.M. & Mitchell, M.J. (2003) Chemical Response of Lakes in the Adirondack Region of New York to Declines in Acidic Deposition. *Environmental Science & Technology*, **37**, 2036–2042.

Fee, E.J., Hecky, R.E., Kasian, S.E.M. & Cruikshank, D.R. (1996) Effects of lake size, water clarity, and climatic variability on mixing depths in Canadian Shield lakes. *Limnology and Oceanography*, **41**, 912–920.

Fergus, C.E., Finley, A.O., Soranno, P.A. & Wagner, T. (2016) LAGOS - Chlorophyll, TP, and water color summer epilimnetic concentrations and lake and catchment data for inland lakes in WI, MI, NY, and ME - a subset of lake data from LAGOSLimno v.1.040.1.

Fergus, C.E., Soranno, P.A., Cheruvelil, K.S. & Bremigan, M.T. (2011) Multiscale landscape and wetland drivers of lake total phosphorus and water color. *Limnology and Oceanography*, **56**, 2127–2146.

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**, 1691–1703.

Finley, A.O. (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, **2**, 143–154.

Gelfand, A.E., Kim, H.-J., Sirmans, C.F. & Banerjee, S. (2003) Spatial Modeling With Spatially Varying Coefficient Processes. *Journal of the American Statistical Association*, **98**, 387–396.

Gelfand, A.E., Schmidt, A.M., Banerjee, S. & Sirmans, C.F. (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, **13**, 263–312.

Gneiting, T. & Raftery, A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Håkanson, L., Blenckner, T., Bryhn, A.C. & Hellström, S.-S. (2005) The influence of calcium on the chlorophyll–phosphorus relationship and lake Secchi depths. *Hydrobiologia*, **537**, 111–123.

Heathcote, A.J., del Giorgio, P.A. & Prairie, Y.T. (2015) Predicting bathymetric features of lakes from the topography of their surrounding landscape. *Canadian Journal of Fisheries and Aquatic Sciences*, **72**, 643–650.

Heathwaite, A.L., Johnes, P.J. & Peters, N.E. (1996) Trends in nutrients. *Hydrological Processes*, **10**, 263–293.

Higgins, J.V., Bryer, M.T., Khoury, M.L. & Fitzhugh, T.W. (2005) A Freshwater Classification Approach for Biodiversity Conservation Planning. *Conservation Biology*, **19**, 432–445.

Jones, R.I. (1992) The influence of humic substances on lacustrine planktonic food chains. *Hydrobiologia*, **229**, 73–91.

Kalff, J.A. (2002) *Limnology: Inland water ecosystems*, Prentice Hall.

Karlsson, J., Byström, P., Ask, J., Ask, P., Persson, L. & Jansson, M. (2009) Light limitation of nutrient-poor lake ecosystems. *Nature*, **460**, 506–509.

Kissman, C.E.H., Williamson, C.E., Rose, K.C. & Saros, J.E. (2013) Response of phytoplankton in an alpine lake to inputs of dissolved organic matter through nutrient enrichment and trophic forcing. *Limnology and Oceanography*, **58**, 867–880.

Kling, G.W., Kipphut, G.W., Miller, M.M. & O'Brien, W.J. (2000) Integration of lakes and streams in a landscape perspective: the importance of material processing on spatial patterns and temporal coherence. *Freshwater Biology*, **43**, 477–497.

Klug, J.L. (2002) Positive and negative effects of allochthonous dissolved organic matter and inorganic nutrients on phytoplankton growth. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 85–95.

Kortelainen, P. (1993) Content of Total Organic Carbon in Finnish Lakes and Its Relationship to Catchment Characteristics. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 1477–1483.

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**, 4425–4435.

Laudon, H., Berggren, M., Ågren, A., Buffam, I., Bishop, K., Grabs, T., Jansson, M. & Köhler, S. (2011) Patterns and Dynamics of Dissolved Organic Carbon (DOC) in Boreal Streams: The Role of Processes, Connectivity, and Scaling. *Ecosystems*, **14**, 880–893.

Mazumder, A. (1994) Phosphorus–Chlorophyll Relationships under Contrasting Herbivory and Thermal Stratification: Predictions and Patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 390–400.

Nürnberg, G.K. & Shaw, M. (1998) Productivity of clear and humic lakes: nutrients, phytoplankton, bacteria. *Hydrobiologia*, **382**, 97–112.

Pan, D., Domon, G., Blois, S. de & Bouchard, A. (1999) Temporal (1958–1993) and spatial patterns of land use changes in Haut-Saint-Laurent (Quebec, Canada) and their relation to landscape physical attributes. *Landscape Ecology*, **14**, 35–52.

Peters, D.P.C., Pielke, R.A., Bestelmeyer, B.T., Allen, C.D., Munson-McGee, S. & Havstad, K.M. (2004) Cross-scale interactions, nonlinearities, and forecasting catastrophic events. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15130–15135.

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

Prairie, Y.T., Duarte, C.M. & Kalff, J. (1989) Unifying Nutrient–Chlorophyll Relationships in Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 1176–1182.

Puckett, L.J. (1995) Identifying the Major Sources of Nutrient Water Pollution. *Environmental Science & Technology*, **29**, 408A–414A.

Rasmussen, J.B., Godbout, L. & Schallenberg, M. (1989a) The humic content of lake water and its relationship to watershed and lake morphometry. *Limnology and Oceanography*, **34**, 1336–1343.

Rasmussen, J.B., Godbout, L. & Schallenberg, M. (1989b) The humic content of lake water and its relationship to watershed and lake morphometry. *Limnology and Oceanography*, **34**, 1336–1343.

Seekell, D.A., Lapierre, J.-F., Ask, J., Bergström, A.-K., Deininger, A., Rodríguez, P. & Karlsson, J. (2015a) The influence of dissolved organic carbon on primary production in northern lakes. *Limnology and Oceanography*, n/a–n/a.

Seekell, D.A., Lapierre, J.-F. & Karlsson, J. (2015b) Trade-offs between light and nutrient availability across gradients of dissolved organic carbon concentration in Swedish lakes: Implications for patterns in primary production. *Canadian Journal of Fisheries and Aquatic Sciences*.

Seekell, D.A., Lapierre, J.-F., Pace, M.L., Gudasz, C., Sobek, S. & Tranvik, L.J. (2014) Regional-scale variation of dissolved organic carbon concentrations in Swedish lakes. *Limnology and Oceanography*, **59**, 1612–1620.

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

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

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

Soranno, P.A., Cheruvelil, 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**, 440–454.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–639.

Taranu, Z.E. & Gregory-Eaves, I. (2008) Quantifying Relationships Among Phosphorus, Agriculture, and Lake Depth at an Inter-Regional Scale. *Ecosystems*, **11**, 715–725.

Thrane, J.-E., Hessen, D.O. & Andersen, T. (2014) The Absorption of Light in Lakes: Negative Impact of Dissolved Organic Carbon on Primary Productivity. *Ecosystems*, **17**, 1040–1052.

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

Webster, K.E., Soranno, P.A., Cheruvelil, K.S., Bremigan, M.T., Downing, J.A., Vaux, P.D., Asplund, T.R., Bacon, L.C. & Connor, J. (2008) An empirical evaluation of the nutrient-color paradigm for lakes. *Limnology and Oceanography*, **53**, 1137–1148.

Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A. & Wade, T.G. (2013) Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment*, **130**, 294–304.

Williamson, C.E., Morris, D.P., Pace, M.L. & Olson, O.G. (1999) Dissolved organic carbon and nutrients as regulators of lake ecosystems: Resurrection of a more integrated paradigm. *Limnology and Oceanography*, **44**, 795–803.

Wu, J. (2004) Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology*, **19**, 125–138.

Xenopoulos, M.A., Lodge, D.M., Frentress, J., Kreps, T.A., Bridgham, S.D., Grossman, E. & Jackson, C.J. (2003) Regional comparisons of watershed determinants of dissolved organic carbon in temperate lakes from the Upper Great Lakes region and selected regions globally. *Limnology and Oceanography*, **48**, 2321–2334.

Yuan, L.L. & Pollard, A.I. (2014) Classifying lakes to improve precision of nutrient–chlorophyll relationships. *Freshwater Science*, **33**, 1184–1194.

Chapter 2: The freshwater landscape: Lake, wetland, and stream abundance and connectivity at macroscales

Abstract

The abundance of and the hydrological connections among lakes, wetlands, and streams can influence broad-scale phenomena such as species dispersion and regional nutrient and carbon processing; and both are likely important for understanding freshwater and terrestrial macrosystem processes at regional to global scales. However, unlike relatively well-studied terrestrial landscapes, scientists have a poor understanding of regional to continental patterns of the surface freshwater landscape that includes lakes, wetlands, and streams together and their connections. In fact, freshwater abundance and connectivity measurements are lacking at macroscales. We measured lake, wetland, and stream abundance and connectivity at a subcontinental extent in the Midwest and Northeast U.S. to study macroscale patterns of the freshwater landscape. We found patterns in both abundance and connectivity of freshwater systems that were related to hydrogeomorphic, climate, and land use variables. Our study describes a robust approach to quantitatively measure freshwater abundance and connectivity, which is needed to integrate freshwater systems into macroscale biogeochemical budgets and models, as well as to inform biodiversity and conservation studies at a variety of scales. Our results also provide insight about potential drivers of freshwater distribution in other regions and continents across freshwater types.

Introduction

Broad-scale disturbances such as land use conversion and climate change are currently altering hydrologic properties at multiple spatial and temporal scales and subsequently threaten the integrity and function of freshwater systems in the landscape (Carpenter *et al.*, 2011). However, the effects of land use disturbance and climate change on freshwater systems are likely to vary across the landscape depending on underlying hydrogeomorphic characteristics (Webster *et al.*, 2000), the spatial configuration of freshwater features (Vörösmarty *et al.*, 2010), and interactions with natural features and human hydrological modifications (Jones *et al.*, 2012). Thus, it is challenging to assess the impacts of these potential disturbances without recognizing the freshwater landscape within the context of its geographic setting and at the broad spatial scales that are aligned with the above scales of disturbance (Jones, 2011; Moore, 2015).

A necessary step before assessing the effects of broad-scale disturbances on freshwaters is to first determine the distribution of freshwater systems in the landscape. Recent progress has been made in estimating regional to global abundance of individual freshwater systems including lakes (Downing *et al.*, 2006; McDonald *et al.*, 2012; Verpoorter *et al.*, 2014), wetlands (e.g., Aselmann & Crutzen, 1989; Lehner & Döll, 2004), and streams (e.g, Downing *et al.*, 2012). This work has been largely motivated by the need to integrate freshwater ecosystems into macroscale carbon and nutrient cycles (Downing *et al.*, 2006; Verpoorter *et al.*, 2014). However, these abundance estimates and inventories are generally conducted on single freshwater system types and do not provide a picture of the integrated surface freshwater landscape that includes lakes, wetlands, and streams, which are active components of biogeochemical processes (but see Raymond *et al.*, 2013; Butman *et al.*, 2016). In addition, such analyses estimate the number and size of freshwater bodies, but they do not calculate the measures of freshwater connectivity that are a critical component of freshwater ecosystem functions. For example, the hydrologic connections (or lack thereof) among lakes, wetlands, and streams are related to variation in chemical and biological attributes (Soranno *et al.*, 1999; Kling *et al.*, 2000), significant differences in carbon, nutrient, and water processes (Cardille *et al.*, 2007; Acuña & Tockner, 2010; Racchetti *et al.*, 2010; Yuan *et al.*, 2015), and the dispersal and movement of organisms influencing meta-population and community dynamics (Pringle, 2001; Crump *et al.*, 2007; Bouvier *et al.*, 2009). Thus, at regional to global scales, scientists currently have an incomplete view of the freshwater landscape that ignores the potential richness and diversity in freshwater abundance and connectivity, which limits their ability to define the role of freshwater ecosystems in regional and global processes.

Until recently, it has been challenging to incorporate freshwater connectivity characteristics into broad-scale frameworks because of computational limitations, a lack of integrated lake, wetland, and stream datasets, and a scarcity of integrated freshwater connectivity landscape measures. Technological advances and national-scale geographic data resources (e.g., the high-resolution U.S. National Hydrology Dataset – NHD) have ameliorated some of these constraints. In addition, landscape measures of freshwater connectivity have been developed for individual freshwater systems that can be applied to broad measures of surface freshwater connectivity. For example, there are several landscape position metrics that capture aspects of lake connectivity related to groundwater flow, surface stream networks, and upstream lakes (Kratz *et al.*, 1997; Soranno *et al.*, 1999; Riera *et al.*, 2000; Martin & Soranno, 2006; Müller *et al.*, 2013). Wetland connectivity metrics have been developed that characterize wetland patches based on spatial configuration with nearby hydrologic features (i.e., lakes and streams) (Weller *et al.*, 1996; Johnson *et al.*, 1997; Devito *et al.*, 2000; Bouvier *et al.*, 2009). Finally, streams are commonly characterized by their position within the stream network, with Strahler stream order being a common measure (Strahler, 1957). These metrics have been related to biogeochemical variables (Lohse *et al.*, 2009; Humborg *et al.*, 2010; Racchetti *et al.*, 2010; Sadro *et al.*, 2011), hydrogeomorphic and limnological characteristics (Martin & Soranno, 2006; Butman & Raymond, 2011; Read *et al.*, 2015), and responses to land use disturbances (Detenbeck *et al.*, 1993; Freeman *et al.*, 2007; Soranno *et al.*, 2015b), demonstrating that they capture spatial characteristics that are relevant to ecological processes. The utility of such position metrics are that they provide landscape-scale estimates of freshwater connectivity that can be directly applied or easily modified to different geographic settings. However, most studies estimate freshwater connectivity within individual catchments or regions and are rarely applied across broad geographic extents that span multiple regions. In addition, these metrics are not commonly quantified or modeled together to explore freshwater connectivity in the landscape across systems (i.e., lakes, wetlands, and streams) (but see, Kling *et al.*, 2000; Cardille *et al.*, 2007; Jones, 2010). Thus the potential for such metrics to contribute to knowledge of freshwaters at macroscales has not been fully realized.

In this study, we seek to describe the macroscale patterns of the freshwater landscape that includes lake, wetland, and stream abundance and their connections. We define freshwater connectivity as the permanent surface hydrologic connections that link lakes, wetlands, and streams, and we measure aspects of freshwater connectivity that are associated with landscape position based on spatial characteristics with surface stream networks. Our objectives were to 1) measure and describe the macroscale patterns of surface freshwater characteristics that included a) abundance, b) connectivity, and c) both abundance and connectivity at a sub-continental spatial extent and 2) relate these macroscale patterns to hydrogeomorphic, land use, and climatic variables that are hypothesized variables associated with the distribution and spatial characteristics of surface freshwaters. We addressed objective one with three different analyses: a) we measured freshwater abundance individually for lakes, wetlands, and streams, b) we measured freshwater connectivity as defined above individually for lakes, wetlands, and streams, and c) we combined measures of both abundance and connectivity that were integrated across lakes, wetlands, and streams together.

Our analyses focused on macroscale patterns of inland freshwater bodies and surface connections and did not include information on very small systems (i.e., lakes smaller than 0.04 km^2), ephemeral systems (i.e., intermittent streams), or groundwater connections. Although we recognize that that such connections play important roles in physical, chemical, and biological processes, high-resolution spatial and temporal data are not available at these broad geographic extents. In addition, permanent surface freshwater connections that we quantify here are major drivers of the movement of materials and organisms between landscape elements.

Methods

Study extent

We quantified a variety of freshwater metrics (described below) for a subcontinental extent (~1,800,000 km^2) in the temperate Midwest and Northeast regions of the U.S. (Fig. 5). The spatial extent is rich in surface freshwater systems (lakes, wetlands, and streams) and spans a wide gradient in hydrogeomorphic, land use, and climatic conditions to capture the diversity of freshwater systems and connectivity characteristics across regional settings.

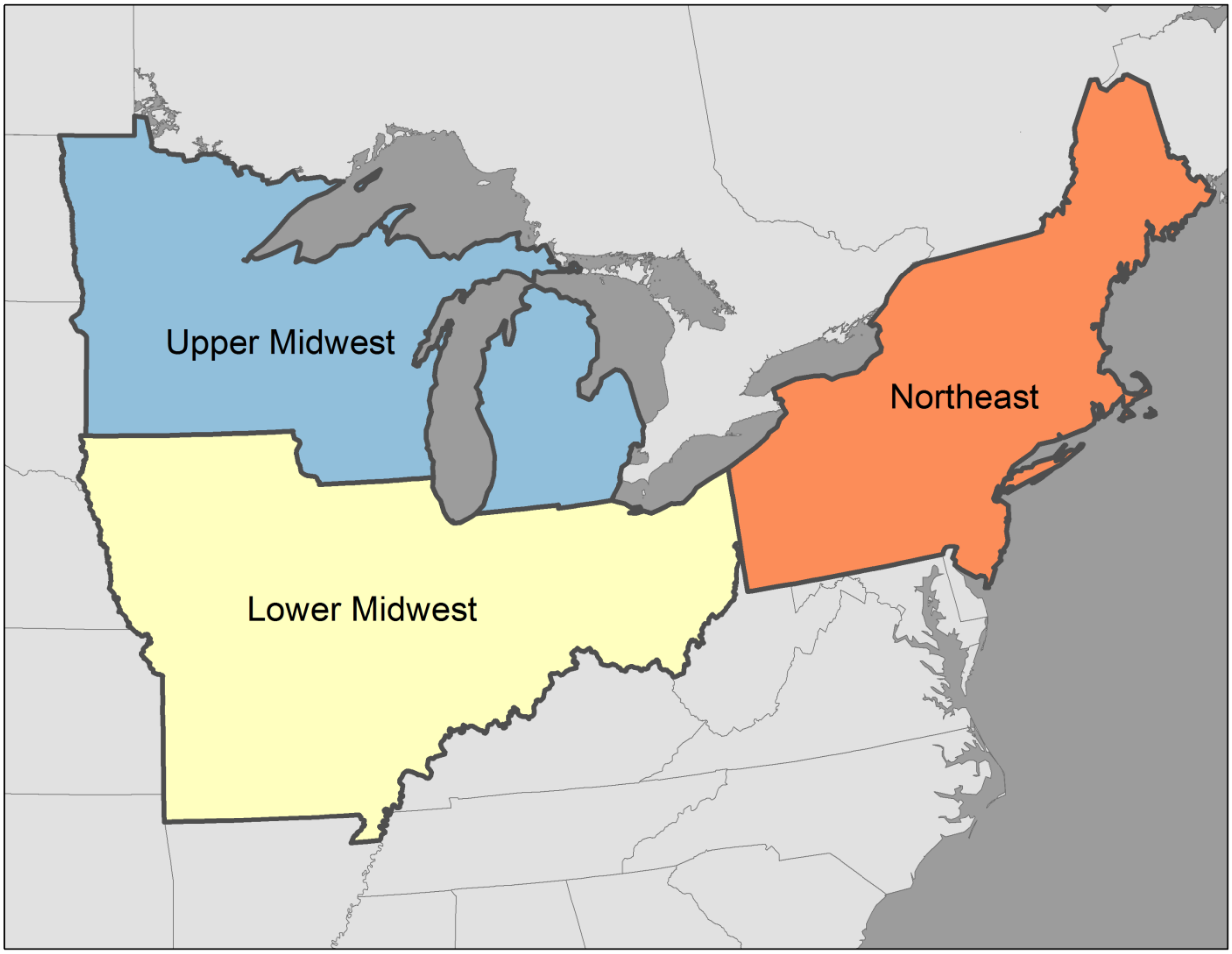

Figure 5. Study extent regions. The study extent spans three regions in the U.S.: the Upper Midwest (Minnesota, Wisconsin, and Michigan), the Lower Midwest (Iowa, Missouri, Illinois, Indiana, and Ohio), and the Northeast (Pennsylvania, New York, New Jersey, Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, and Maine).

Description of datasets

The data used in this study come from the LAGOS database (Lake multi-scaled geospatial and temporal database (Soranno *et al.*, 2015a). We used $LAGOS_{GEO}$, version 1.03, an integrated, multi-thematic geographical database that includes national-scale data for geology, topography, hydrology, climate, and land use/land cover. Lake and stream data in $LAGOS_{GEO}$ version1.03 originally came from the National Hydrography Dataset (NHD – United States Geological Survey – Version 9.3; 1:24,000 resolution), and wetland data come from the National Wetlands

Inventory (NWI – U.S. Fish and Wildlife Service). See Soranno et al. (2015a) for download dates for each of these datasets.

Additional geographic data in $LAGOS_{GEO}$ version 1.03 came from other data sources with full citations for each dataset provided in Soranno et al. (2015a). Regional hydrology data came from the National Water Information System portal (USGS). Topographic data came from the National Elevation Dataset (USGS). Geology data and glaciation limits from the Wisconsin glacial period (the most recent glaciation event in North American, ~2.6 million to 11,000 years ago) came from surficial materials map database (USGS). Climate data come from the PRISM Climate Group – 30 year normal data. Land use/land cover data come from the 2011 National Land Cover Database (Multi-Resolution Land Characteristic Consortium).

Definitions of lakes, wetlands, and streams

Prior to quantifying freshwater abundance and connectivity, we restricted analysis of freshwater geographic features (lakes, streams, and wetlands) based on accuracy of the geographic data to capture freshwater systems and relevant attributes. This was a necessary step because both the NHD and NWI comprehensively map all freshwater feature types including artificial systems (e.g., sewage treatment ponds) and artificial flowlines (e.g., lines connecting NHD features) that we did not consider to be part of the surface freshwater landscape.

We defined lakes as perennial water bodies ≥ 0.04 km^2, including both natural lakes and reservoirs (i.e., impounded streams or rivers). Lakes smaller than 0.04 km^2 in size were excluded from the analyses because these smaller water bodies were associated with high identification and digitization error rates compared to lakes that were 0.04 km^2 and larger in the NHD data layer (for error rate analysis see Soranno *et al.*, 2015). We did not distinguish between natural lakes and reservoirs because the NHD identifies lakes as reservoirs only for water bodies that could be easily identified as artificial and many artificial systems are mislabeled as lakes (McDonald *et al.*, 2012). Thus, we currently lack high-resolution, integrated dam and lake data to accurately differentiate system types especially in highly human-modified landscapes (e.g., agricultural regions).

Wetland data included NWI classified *Palustrine* systems – i.e., non-tidal wetlands dominated by trees, shrubs, and persistent emergent vegetation. NWI reliably captures wetlands that are 0.002 km^2 in size or larger with a 98% identification accuracy rate (Wetland Mapping Standard — Federal Geographic Data Committee).

Finally, stream data included perennial streams classified as *Stream/River* and *Canal/Ditch* feature types in the NHDFlowline data layer. Freshwater feature definitions and detailed descriptions of preprocessing steps for the data in $LAGOS_{GEO}$ version 1.03 are provided in Soranno *et al.*, 2015a.

Freshwater abundance and connectivity metrics

We measured freshwater abundance and connectivity metrics at two spatial scales – Hydrologic Units (HU)12 and HU8. The HUs are hierarchically nested stream watershed spatial units that are based on USGS 1:24,000 scale topographic maps. The HU12 scale is the smallest nested HU in $LAGOS_{GEO}$ and was used to examine fine-scale heterogeneity in freshwater attributes (n = 18870; median size = 78.10 km^2). The HU8 scale is an intermediate sized spatial extent and was used to examine regional heterogeneity in freshwater attributes (n = 445; median size = 2880 km^2). $LAGOS_{GEO}$ version 1.03 records the summed surface area of lakes and wetlands and stream length and count of lake and wetland polygons (total and by connectivity type) within HUs. Freshwater connectivity metrics in $LAGOS_{GEO}$ version1.03 were calculated by first assigning lake, wetland, and stream features into connectivity groups based on their spatial arrangement with other freshwater features and then measuring surface area or stream length of each connectivity group within the spatial unit. Using these values, we analyzed freshwater abundance and connectivity across the study extent.

Freshwater abundance was measured as the total proportion area (lake or wetland) or total length density (stream) within a spatial unit. Freshwater connectivity was calculated by taking the relative proportion of the connectivity type out of the total area or length of the respective freshwater system (e.g., isolated lake area divided by total lake area within the HU12). Relative proportion metrics are better suited to our analyses compared to total areal proportion metrics (e.g., isolated lake area divided by HU12 area) because freshwater features cover only a small fraction of area across our macroscale extent, resulting in many observations having zero or very low values. Thus, relative proportions can capture the relative distribution of freshwater connectivity groups where freshwater features are present on the landscape. The specific metrics are described below and illustrated in Fig. 6. Detailed descriptions of the geoprocessing steps and the GIS toolbox developed to calculate the connectivity metrics are described in Soranno *et al.*, (2015a) and available at https://github.com/soranno/LAGOS_GIS_Toolbox.

<u>Lake connectivity types</u> – Lake connectivity types captured aspects of the landscape position of the lake – the surface hydrologic connections of the focal lake with inflowing and outflowing streams and upstream lakes. Lakes were grouped into four hydrologic classes: Isolated $_{\text{lake}}$, Headwater $_{\text{lake}}$, Drainage $_{\text{lake}}$, and Drainage-UPLK $_{\text{lake}}$ (Fig. 6). Isolated lakes have no stream inlets or outlets. Headwater lakes have no stream inlet and at least one outlet. Drainage lakes have inlets and outlets and no upstream lakes (≥ 0.10 km^2 in size). Drainage-UPLK lakes have inlets and outlets and at least one upstream lake (≥ 0.10 km^2 in size).

Wetland connectivity types – Wetland connectivity types also were measures of wetland landscape position with surface stream networks. Streamflow direction was not included in these metrics because of a lack of computing power to make such calculations on the large number of wetland polygons in the NWI (~ millions of polygons). Wetlands were grouped by surface hydrologic relationships with stream segments into three classes: Isolated $_{wetland}$, Headwater $_{wetland}$, Drainage $_{wetland}$ (Fig. 6). Wetland connectivity to streams were determined if wetlands were within a 30 m buffer surrounding the stream reach to accommodate spatial data resolution limitations and misalignment between different data layers (NHD and NWI). Isolated wetlands have no stream inlets and outlets. Headwater wetlands were wetlands intersected by a Strahler first order stream segment. Drainage wetlands were wetlands intersected by either a single >1st order stream segment or multiple stream segments.

Stream connectivity types – Stream features in our study extent were assigned connectivity groups based on Strahler stream order. Strahler stream order is a method to characterize the position of the stream reach in relation to the stream network with stream order increasing as one moves from the headwater streams to the terminal point (Strahler, 1957). This relatively simple metric has been associated with stream attributes such as catchment size and mean annual discharge (Hughes *et al.*, 2011) and processes such as CO_2 flux (Butman & Raymond, 2011) indicating that is a relevant metric that captures aspects of stream hydrology and ecology. Stream order was calculated using RivEx 10.6 GIS tool (Hornby, 2010). Stream order classes were grouped into three categories: Low-order, Mid-order, and High-order (Fig. 6). Low-order streams include 1st – 3rd order stream reaches, Mid-order streams include 4th – 6th order reaches, and High-order streams include greater than 6th order reaches.

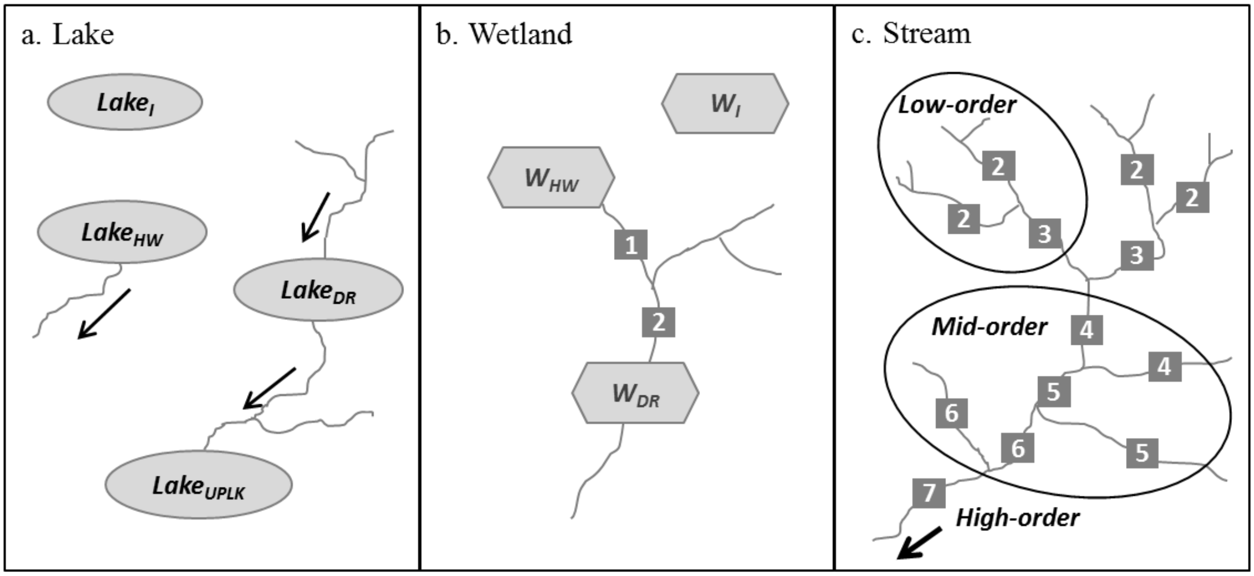

Figure 6. Diagram of freshwater connectivity metrics for a) lakes, b) wetlands, and c) streams. Lakes (oval) were assigned connectivity types based on spatial relationships with streams and upstream lakes. Wetlands (hexagon) were assigned connectivity types based on spatial relationships with streams. Lake connectivity types include Isolated ($Lake_I$) – no stream inlets or outlets, Headwater ($Lake_{HW}$) – only stream outlets, Drainage ($Lake_{DR}$) – stream inlets and outlets, and Drainage-UPLK ($Lake_{UPLK}$) – stream inlets and outlets and at least one upstream lake ≥ 10 ha in size. Wetland connectivity types include Isolated (W_I) – no intersecting stream segments, Headwater (W_{HW}) – intersected by one first order stream segment, and Drainage (W_{DR}) – intersected by a stream segment > 1^{st} order or multiple stream segments. Stream segments were assigned Strahler stream order (squares) based on location in the stream network and grouped into stream connectivity classes: Low-order (1^{st} – 3^{rd}) , Mid-order (4^{th} – 6^{th}) , and High-order (> 6^{th}).

Freshwater abundance and connectivity patterns

We performed separate analyses on freshwater abundance and connectivity for each freshwater system type (i.e., lakes, wetlands, and streams) at the HU12 scale, and we performed analyses integrating freshwater abundance and connectivity across all three freshwater system types at the HU8 scale.

Patterns of surface freshwater abundances – We quantified freshwater abundance (i.e., total proportion or density) at the HU12 scale and mapped quartile values (25^{th} percentile, 50^{th} percentile, and 75^{th} percentile) for lakes, wetlands, and streams separately.

Patterns of surface freshwater connectivity – Freshwater connectivity patterns were analyzed separately for lake, wetland, and stream connectivity metrics at the HU12 scale, and jointly at the HU8 scale. We used a combination of principal components and k-means cluster analyses to examine freshwater connectivity. Prior to analyses, we transformed the relative proportion (logit) and stream density (natural log) values to meet assumptions of normality and homoscedasticity. Principal component analyses (PCA) were performed on the transformed freshwater connectivity metrics to remove codependence among metric values (JMP software). Cascade K-means cluster analyses were preformed to group HU12s in our study extent that had similar freshwater connectivity characteristics and to differentiate HU12s with dissimilar connectivity characteristics (JMP software). We determined the number of cluster groups based on the eigenvalues from the kernel matrix to optimize the compactness and accuracy in the cluster groups. There were some spatial units that were not assigned cluster groups because they were no freshwater systems present (lakes; n = 7922) or because stream segments were not assigned Strahler stream order class due to digitization errors in the original NHD data (n = 123). We then mapped freshwater connectivity cluster groups across the study extent to visualize the patterns of surface freshwater connectivity at macroscales. The maps depicted freshwater connectivity cluster groups by the dominant connectivity type present (based on the PCA scores).

Patterns of abundance and connectivity integrated across lakes, wetlands, and streams – We performed a similar set of analyses as above (i.e., PCA and cluster analyses) to examine both freshwater abundance and connectivity patterns together among lakes, wetlands, and streams (i.e., integrated freshwater abundance and connectivity). However, we analyzed all freshwater metrics together using data quantified at the HU8 scale. The HU8 scale was used rather than the HU12 scale because there were many HU12s with zero lakes present resulting in a highly skewed distribution that would violate statistical assumptions of normality in the PCA and k-means cluster analyses. The HU8 scale, however, spanned a large enough area to produce normal distributions in the freshwater metrics.

Hydrogeomorphic, climate, and land use variables associated with freshwater abundance and connectivity

To address objective two, we performed random forest analyses to examine whether the integrated freshwater cluster groups (at HU8 scale) were associated with underlying landscape and climatic characteristics (Appendix Table B1). Random forest is a machine-learning technique based on classification regression trees and combines multiple classification trees to improve classification accuracy (Cutler *et al.*, 2007). The algorithm uses bootstrap samples of the original observations and randomized subsets of predictor variables to build individual trees. Model accuracy and variable importance were estimated from the hold-out observations to evaluate the association between landscape and climatic variables and integrated freshwater cluster groups.

Results and Discussion

Our analyses of lake, wetland, and stream abundances and surface connectivity present a synthetic view of the freshwater landscape, which provides a comprehensive view of the diversity of freshwater systems at macroscales. We found that (1) freshwater abundance of lakes and wetlands exhibited inverse spatial patterns compared to stream density that followed glaciation extent boundaries. Lake and wetland abundance was higher in glaciated areas compared to unglaciated areas; whereas stream density was lower in glaciated areas compared to unglaciated areas. In addition, we found (2) distinct, broad-scale patterns among lake, wetland, and stream connectivity that reveal a layer of complexity that abundance measures alone did not capture, suggesting that freshwater abundance and connectivity may be influenced by different underlying processes. And, (3) abundance and connectivity measures integrated across all three freshwater system types (i.e., lakes, wetlands, and streams) showed spatially contiguous patterns; and (4) these patterns were associated with underlying hydrogeomorphic, climate, and land use variables. We illustrate a robust approach to quantitatively measure freshwater abundance and connectivity to incorporate into models at macroscales. These results can inform both fundamental and applied scientific questions to better understand and manage freshwater systems.

(1) Patterns of surface freshwater abundances for lakes, wetlands, and streams

Freshwater abundance estimates are typically conducted at regional to global scales in other studies, and our results were in agreement with the spatial patterns and range of values reported in the literature. We found that lake, wetland, and stream abundances, measured as areal proportions or stream density within the HU12, exhibited different macroscale patterns within the study extent (Fig. 7 a – c). Lakes (≥ 0.04 km^2 in size) were only present in a little over half of the study extent (58% of HU12 spatial units) in contrast to wetlands and streams that were found throughout the study extent (~99% of the HU12 spatial units).

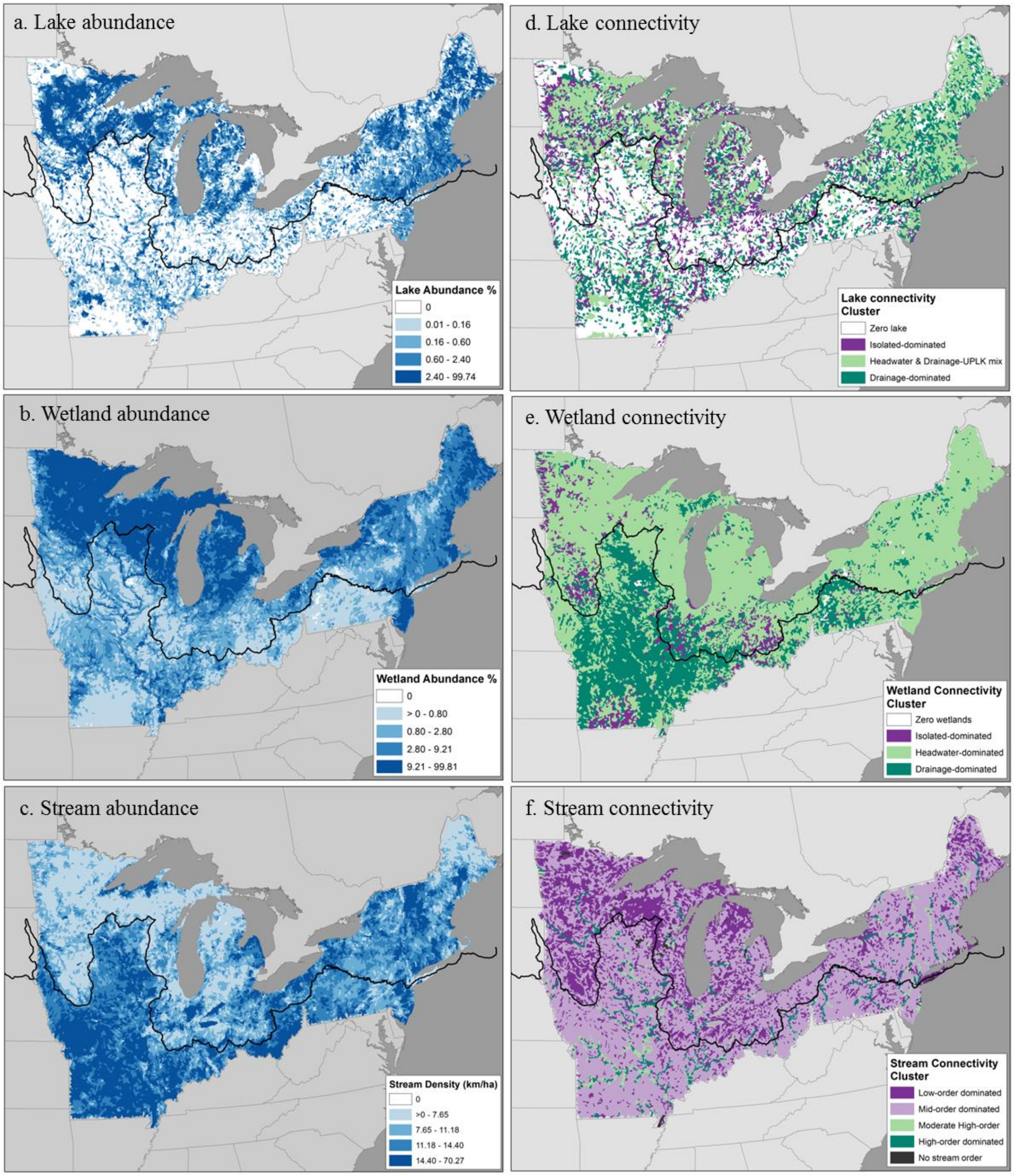

Figure 7. Freshwater abundance and connectivity maps by system type. Freshwater abundance is quantified as the total proportion area or length within the HU12 spatial unit for lakes (a), wetlands (b), and streams (c). Abundance values are binned as quartiles. Freshwater connectivity for lakes (d), wetlands (e), and streams (f) is represented by connectivity cluster groups determined by k-means cluster group using PCA scores from lake, wetland, and stream connectivity metrics. The solid black line represents the estimated boundary of the Wisconsin glacial period – north of the line is glaciated and south of the line is unglaciated.

In total, lake area made up 1.9% of the total study extent area (with mean HU12 areal percent of 2.2% in glaciated and 0.5% in non-glaciated regions), and lake density was 0.03 km^{-2} (Table 4). Wetlands, on the other hand, made up 7.8% of the total study extent area and were found at a higher density of 2.78 km^{-2}. Total stream length in the study extent was 1,794,044 km with a total stream density of 1.13 km/km^2. Our mean lake areal proportion estimates (HU12) are similar to estimates for the entire U.S. described by Meybeck (1995; 2.8% glaciated and 0.09% non-glaciated). Our wetland areal proportion estimate of 7.8% is slightly higher than the estimate of 5.5% by Dahl (2005) for the conterminous U.S., but this discrepancy is not surprising because our study extent did not included arid and mountainous regions that have low wetland abundance. In addition, we found that lake density (0.03 km^{-2}) was lower than wetland density (2.8 km^{-2}) as might be expected given size-frequency relationships that characterize freshwater abundance at macroscales (e.g., Downing *et al.*, 2006) and wetlands being on average smaller than lakes with mean surface area of 0.03 km^2 compared to lake mean surface area of 0.6 km^2.

In general all three ecosystem abundance spatial patterns followed the Wisconsin glaciation boundary but with inverse patterns for lakes and wetlands compared to streams. In the study extent, the majority of spatial units were glaciated; ~ 60% of the HU12s were categorized as Glaciated, 37% were categorized as Unglaciated, and 3% were categorized as Partially Glaciated. Lakes were largely absent south of the glaciation line but were also missing in glaciated parts of the Lower Midwest region (Fig. 7a). Wetland abundance generally was higher north of the glaciation boundary compared to the south (Fig. 7b; Appendix Fig. B2). Stream density was more irregular in its patterns. For example, in the Upper Midwest region, stream abundance was higher south of the glaciation boundary and lower in the north for the Upper Midwest region. However, this pattern did not hold in the Northeast region where stream density was high both north and south of the glaciation boundary (Fig. 7c). Other studies support that lake abundance is greater in glaciated areas compared to non-glaciated areas for large lakes (≥ 1 km^2) (Meybeck, 1995; Lehner & Döll, 2004). However, there is less information about wetland and stream distributions in relation to glaciation regime.

Table 4. Summary of freshwater abundance and connectivity summarized over the full study extent. Connectivity types are italicized.

Freshwater system	Type	Surface area %	Water body density (km^{-2})	Stream density (m/km^2)
Lake	Total	1.93	0.03	
	Isolated	0.17	0.01	
	Headwater	0.01	0.005	
	Drainage	0.40	0.01	
	Drainage-UPLK	1.2	0.006	
Wetland	Total	7.8	2.78	
	Isolated	2.80	1.90	
	Headwater	0.95	0.28	
	Drainage	4.10	0.64	
Stream	Total			1131
	Low-order			980
	Mid-order			130
	High-order			10

Our lake abundance estimates do not capture very small systems (< 0.04 km^2) nor do they distinguish between natural lakes and reservoirs, which affect the interpretation of the results. For example, other studies have found that small water bodies are found in high densities in unglaciated, agricultural areas in the U.S. (Lower Mississippi region) compared to other parts of the U.S. likely due to the construction of artificial ponds (Smith *et al.*, 2002) and our lake cutoff may exclude very small, artificial waterbodies and subsequently underestimate lake abundances especially in agricultural areas. In fact, a study by McDonald et al. (2012) that used the same

NHD dataset (1:24,000) that we used but that also included lakes ≥ 0.001 km^2 to estimate regional lake abundance reported higher lake areal coverage and density (2.3% and 0. 53 km^{-2}) compare to our estimates (1.9% and 0.03 km^{-2}). It would be expected that extending the dataset to include very small water bodies <0.04 km^2 and ≥ 0.001km^2 would increase the lake number exponentially such that density estimates would be significantly affected by lake size cutoffs more so than areal coverage estimates (McDonald *et al.*, 2012).

Although our freshwater abundance estimates do not capture very small systems, the intent of this study was not to provide an exhaustive inventory across freshwater size classes but rather to develop methods to integrate across freshwater types and combine measures of abundance and connectivity. Small water bodies are important from a hydrological and ecological standpoint and these systems may be very dynamic through time with small artificial ponds being constructed at increasing rates in agricultural regions (Smith *et al.*, 2002) and changes on the landscape potentially increasing the number of temporary, intermittent systems (Datry *et al.*, 2014). However, the spatial and temporal resolution of current national-scale geographic data is not able to accurately capture small and ephemeral freshwater systems. For example, the target mapping unit for wetlands in the NWI data ranged from 0.004 km^2 to 0.01 km^2 (Tiner, 1997) and smaller wetland systems cannot be assessed to a reliable degree. In addition, stream data in the NHD likely underrepresent small, intermittent, and ephemeral reaches (Nadeau & Rains, 2007; Roy *et al.*, 2009). Therefore, there is a need for high resolution spatial and temporal data at broad spatial extents to incorporate small, temporary freshwater systems in macroscale analyses.

An important outcome of our work is the large spatial variability in freshwater abundance at regional to sub-continental scales. However, because freshwater inventories or high resolution geospatial data are often lacking at broad spatial extents, freshwater abundance is sometimes estimated using size-distribution scaling laws to extrapolate to new regions or continents as has been done for lakes (Meybeck, 1995; Downing *et al.*, 2006, 2012; Raymond *et al.*, 2013). But these lake size-distribution relationships can be inconsistent across different regional settings and can lead to erroneous lake abundance estimates if regional differences in abundance are ignored (Seekell & Pace, 2011; Seekell *et al.*, 2013). Stream scaling relationships may be more robust to different regional contexts but require high resolution data to accurately infer stream order (Downing *et al.*, 2012). Thus, our approach of using high resolution geospatial data with good spatial coverage presents a step forward to estimate freshwater abundance at macroscales.

(2) Patterns of surface freshwater connectivity of lakes, wetlands, and streams

Freshwater connectivity revealed heterogeneity in the freshwater landscape that the abundance metrics did not capture. In addition, we found that the distributions and regional patterns were different for lake, wetland, and stream connectivity types (Fig. 7 d – f). Across the entire study extent, lake connectivity types were not uniformly distributed. The differences in freshwater connectivity patterns in contrast to freshwater abundance patterns suggest that freshwater connectivity is somewhat independent of freshwater abundance, and ignoring freshwater connectivity may omit important, distinct hydrologic characteristics.

For the PCA analyses, we retained the first two principal axes that explained high amounts of variation in the metric values (lakes 70.6%, wetlands 96.5%, and streams 99.1%). For the k-means cluster analysis, we identified three connectivity clusters for lake and wetland connectivity composite scores and four connectivity clusters for stream connectivity composite scores that captured the dominant connectivity attribute. Principal component analyses and k-means cluster results are provided in Appendix Fig. B3.

Freshwater connectivity areal proportion values (% area in HU12) followed the same general trends as the total proportions by freshwater system type in relation to glaciation regime (Appendix Fig.B 4). For example, across lake connectivity types, areal proportions were higher in glaciated areas compared to unglaciated areas. However, the relative proportion of freshwater connectivity types deviated from this trend in some cases (Appendix Fig. B5). For Isolated and Headwater lakes, relative proportions were similar across glaciation regimes. However, the relative proportion of Drainage lakes and Drainage wetlands were higher in non-glaciated areas compared to glaciated and partially glaciated areas – suggesting that although Drainage lakes and wetlands made up smaller portions of the surface area in non-glaciated areas, they tended to be the more dominant lake type present compared to in glaciated areas where there was more heterogeneity in lake connectivity types present.

Drainage lakes were found in the majority of HU12s where lakes were present (~70%); but Isolated, Headwater, and Drainage-UPLK lake types were less common and found in only about half of the HU12s with lakes present. The mapped lake connectivity clusters showed spatial variation in what lake connectivity types were most dominant (Fig. 7d). In the Upper Midwest there was a variety of lake connectivity types present with areas dominated by Isolated lakes as well as areas dominated by lake chains (i.e., cluster groups dominated by Headwater and Drainage-UPLK lakes). In contrast, the Northeast region was mainly dominated by both types of connected lake systems with few areas where Isolated lakes dominated. The Lower Midwest had less area where lakes were present and exhibited a mix of lake connectivity types. If lakes smaller than 0.04 km^2 were included in our analysis, the connectivity is likely to have differed, perhaps increasing the number of isolated systems.

Wetland connectivity types were more evenly distributed in the study extent compared to lake connectivity types, with all connectivity types found in over 90% of the study extent in comparison to lake connectivity types (Fig. 7e). In general, areas north of the glaciation boundary were largely dominated by Headwater wetlands (i.e., wetlands connected to a 1st order stream), and areas south of the glaciation boundary were dominated by Drainage wetlands (i.e., wetlands connected to > 1st order streams). Isolated wetlands were prominent in particular areas throughout the study extent both north and south of the glaciation boundary and typically were the dominant connectivity type within a HU12 unit (54%). Areas where isolated wetlands were prominent in our study extent (Fig. 7e) seem to correspond with locations of specific isolated wetland types such as prairie pothole, kettle-hole, Great Lakes Alvars, and karst wetlands (Tiner, 2003).

Isolated lakes and wetlands made up small proportions of their total respective water body areas, but they had the highest densities compared to other connectivity types (Appendix Table B6). These results align with size-frequency findings that small water bodies dominate freshwater distributions in terms of numbers but not necessarily in terms of surface area (Smith *et al.*, 2002; McDonald *et al.*, 2012; Cohen *et al.*, 2016). In our study extent, isolated systems tended to be smaller on average compared to stream-connected systems (Appendix Table B6). This association between freshwater connectivity type and water body size suggests that freshwater connectivity types may follow size-frequency scaling distributions to some degree, which could aid in extrapolation methods to estimate connectivity at macroscales. In fact for wetlands, the probability of a wetland being isolated does increase with decreasing size (Cohen *et al.*, 2016). However, this association has not been formally evaluated with lakes.

Stream connectivity types defined by stream order were not uniformly distributed in the study extent. Low-order streams were the most ubiquitous stream type (present in 99% of HU12s in the study extent) followed by Mid-order streams (73% of the HU12s) and High-order streams (7% of HU12s). Low-order streams were most prominent in the Upper Midwest and the northern tip of the Northeast region (Fig. 7f). Mid-order streams were dominant throughout the Lower Midwest and Northeast regions. Areas dominated by High-order streams followed major river paths such as the Mississippi River and the Connecticut River, validating the freshwater connectivity metrics and our analytical framework to capture relevant freshwater attributes.

Low-order (1^{st}-3^{rd}) streams dominated total stream length (88%). This trend is in agreement with global and regional stream abundance estimates that indicate that low-order streams have the greatest number of stream segments and make up the most total stream length (Butman & Raymond, 2011; Downing *et al.*, 2012). The proportion of stream length declined with increasing stream order class with Mid-order streams making up 11% and High-order streams making up 1% of the total stream length.

Incorporating surface freshwater connectivity attributes in the macroscale analysis revealed layers of complexity of the freshwater landscape that abundance measures alone did not capture. Most regional and global freshwater distribution studies are performed individually on lakes, wetlands, and streams (except see Lehner & Döll, 2004; Aufdenkampe *et al.*, 2011) and largely ignore their relationships with other freshwater systems in the landscape. Here, we measured freshwater systems by their surface connectivity to other freshwater systems and found evidence that freshwater connectivity exhibits distinct, non-random spatial patterns at broad extents that may provide insight on macroscale relationships that influence freshwater distributions and their connections to the surrounding landscape.

(3) The freshwater landscape: Integrating lake, wetland, and stream abundances and connectivity

Overall, integrating freshwater abundance and connectivity measures for all system types (i.e., lakes, wetlands, and streams) at the HU8 scale revealed broad-scale spatial patterns in the freshwater landscape (Fig. 8). HU8 spatial units were grouped into five cluster groups capturing similar lake, wetland, and stream abundance and connectivity characteristics (based on PCA scores; see Appendix Fig. B7). The interpretation of the integrated freshwater cluster groups are provided in Table 5. At the HU8 scale we lose resolution to capture fine scale variation and heterogeneity in the freshwater landscape compared to the HU12 scale, but the overall patterns generally matched the patterns at the HU12 scale. We found that areas north of the glaciation extent were dominated by lakes and wetlands and were composed of a variety of both isolated and stream-connected system types; and areas south of the glaciation extent were dominated by streams and stream-connected lake and wetland connectivity types. The somewhat contiguous spatial patterns of the freshwater landscape suggest that there may be underlying geographic variables associated with these patterns.

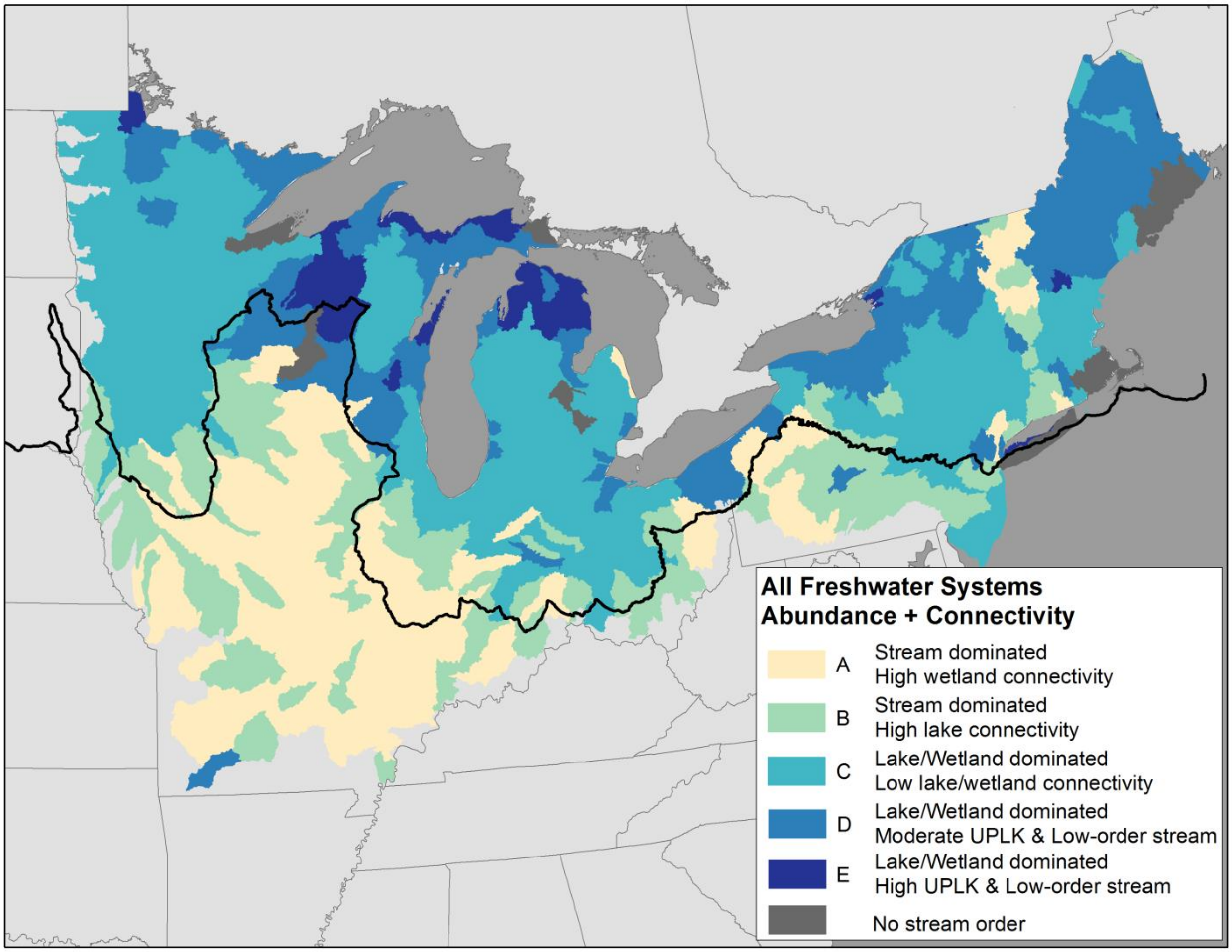

Figure 8. Integrated freshwater abundance and connectivity map. Integrated freshwater abundance and connectivity clusters were assigned using k-means cluster analysis based on PCA scores from lake, wetland, and stream abundance and connectivity metrics quantified at the HU8 scale. Spatial units not assigned a cluster group were missing stream order data for a portion of the area. Interpretation of cluster groups is provided in Table 1. The solid black line represents the estimated boundary of the Wisconsin glacial period – north of the line is glaciated and south of the line is unglaciated.

Table 5. Description of integrated freshwater abundance and connectivity cluster groups at the HU8 scale. Cluster groups were assigned by k-means cluster analysis using PCA scores of lake, wetland, and stream abundance and connectivity metrics.

Cluster assignment	Description/Interpretation
A	High stream density, mix of Low-order $_{stream}$ and High-order $_{stream}$ Low lake and wetland abundance, high proportion Drainage $_{wetland}$
B	High stream density, high proportion Mid-order $_{stream}$ and High-order $_{stream}$ Low lake and wetland abundance, high proportion Drainage $_{lake}$
C	Low stream density; high proportion Mid-order $_{stream}$ High wetland and lake abundance; high proportion Isolated $_{wetland}$ and Headwater $_{wetland}$, and high proportion Isolated $_{lake}$ and Headwater $_{lake}$
D	Low stream density; high proportion Low-order $_{stream}$ High wetland abundance; High lake abundance
E	Low stream density; High lake and wetland abundance, high proportion Drainage-UPLK $_{lake}$

(4) Hydrogeomorphic, climate, and land use variables associated with broad-scale freshwater landscape attributes

We found that integrated freshwater abundance and connectivity clusters were associated with hypothesized hydrogeomorphic, climate, and land use variables with glaciation regime being a top predictor. The associations suggest that past geological activity (i.e., glacial and fluvial processes) may be key drivers that affect the presence of freshwater systems; and hydrologic, geologic, and human land use activity may modify freshwater characteristics across regions in different ways.

The top performing random forest model using hypothesized hydrogeomorphic, climate, and land use variables accurately predicted over 63% of integrated freshwater cluster assignment

(HU8 scale) based on out-of-bag samples. Cluster group C had the highest classification accuracy of 78.5% (21.5% classification error rate) and was characterized as being dominated by lakes and wetlands with low connectivity (Appendix Table B8). Cluster group E had the lowest classification accuracy of 26.1% and was characterized as being dominated by high Drainage-UPLK and Low-order streams. Cluster group E was the least common group with only 23 observations out of 445 HU8s. As expected, the most important variables to predict the integrated clusters were glaciation regime, hydrology (mean runoff and baseflow), geology (glacial fluvial outwash and alluvial deposits), mean precipitation, and human land use (pasture and agriculture) (Appendix Fig.B9). Generally cluster groups A and B were located south of the glaciation boundary and C, D, and E were located north of the glaciation boundary. The relative distributions of important hydrogeomorphic, climate, and land use variables are presented in Fig. 9 by cluster group. In general, areas that were lake and wetland rich (C, D, and E) tended to have high runoff, baseflow, and glacial fluvial outwash geology. In contrast, areas that were rich in streams (A and B) tended to have high precipitation, topographic slope, alluvial geology, and pasture and agricultural land use activities.

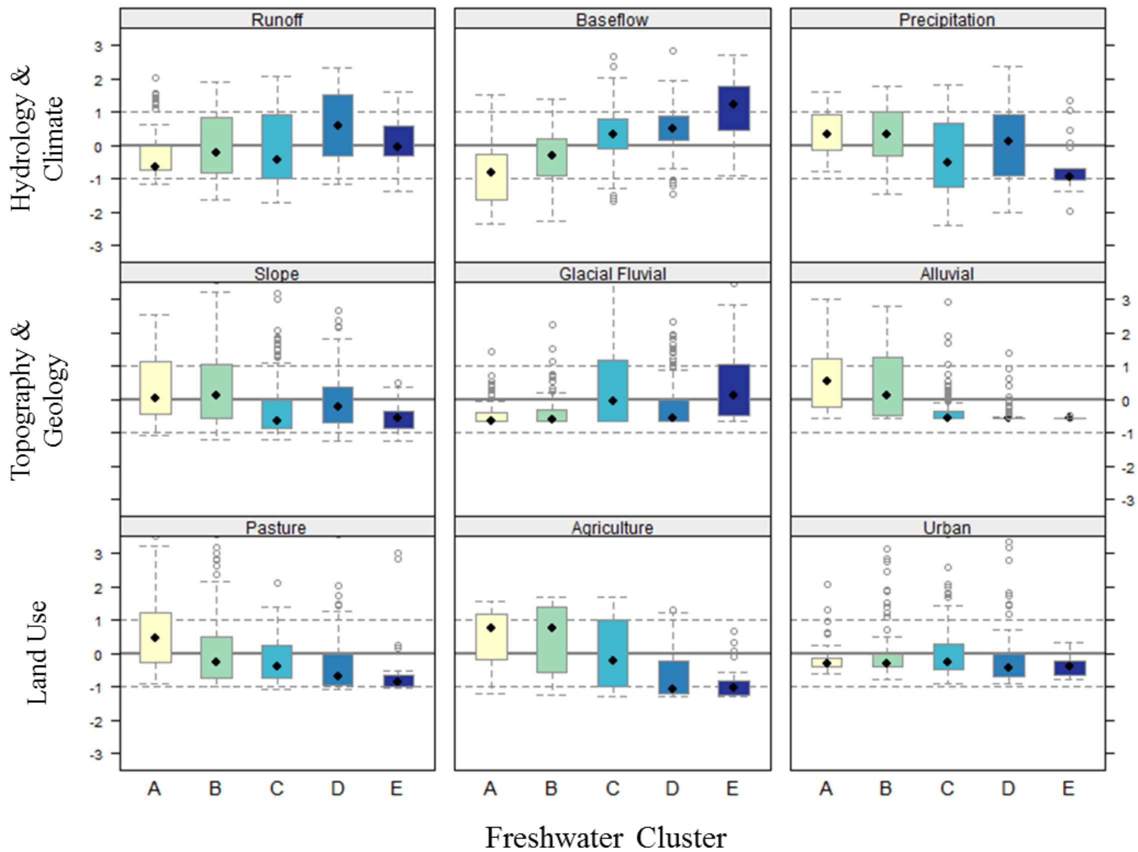

Figure 9. Boxplots of the geophysical predictors of the integrated freshwater abundance and connectivity clusters. Boxplots represent the standardized distributions of geophysical variables among the integrated freshwater HU8 clusters (letters along horizontal axis). Geophysical variables were the top predictors of the integrated freshwater clusters in random forest analyses and are grouped as Hydrology and Climate, Topography and Geology, and Land Use. Values were standardized by subtracting the mean and dividing by the standard deviation. The solid gray line represents the mean and the dashed lines represent ±1 standard deviation above or below the mean.

The composition and connectivity characteristics of the freshwater landscape are shaped by a suite of drivers that influence landscape depressions (past geologic activity, topographic relief), water source (climate and hydrology), permeability of substrate to hold water (surficial geology), and human modifications to the landscape (land use activities). Our exploratory analysis identified hydrogeomorphic, climate, and land use variables that pertain to these different drivers. Glaciation regime was one of the top predictors of the integrated freshwater clusters. The most recent glacial period (Wisconsin stage) scoured the landscape to create depressions and sedimentary deposits, and glaciated areas are associated with increased lake and wetland abundance (Meybeck, 1995; Winter, 2000; Lehner & Döll, 2004). With over half of our study extent being glaciated (59%), it is expected that glaciation regime would be significantly related to patterns in the freshwater landscape and associated with increased lake and wetland abundance.

However, glaciation is not an exclusive driver of the freshwater landscape. Within glaciated areas there can be a great deal of variability in the abundance and types of freshwater bodies present due to differences in rock type and relief (Meybeck, 1995; Winter, 2000). In addition, lakes and wetlands are formed by other processes such as fluvial activity or human hydrologic modifications to the landscape (Meybeck, 1995; Smith *et al.*, 2002). Mean annual runoff and baseflow measures at the HU8 scale were associated with freshwater landscape clusters – with lake- and wetland-rich areas tending to have higher runoff and baseflow compared to areas that were stream-rich. These variables can be indicators of surface and groundwater sources to freshwater systems. Surficial geology was also important to predict freshwater clusters and may be indicators of past geologic activity as well as current hydrologic conditions that influence freshwater presence and composition. High glacial fluvial outwash deposits were associated with lake- and wetland- rich areas and are indicators of past glacial activity (meltwater from retreating glaciers deposit sediment, sand, and gravel) and can be indicative of present groundwater aquifers. Stream-rich areas had high alluvial deposits – clay, silt, sand, and gravel, which are geologic materials associated with past geomorphic processes that shaped the landscape and can characterize current substrate permeability characteristics that influence surface and subsurface water exchange (Winter, 2000; Fisher *et al.*, 2004). Mean precipitation was slightly higher in areas with high stream densities compared to areas with low stream densities. This trend was also observed at a global scale where a positive correlation between stream abundance and precipitation was observed (Raymond *et al.*, 2013). Finally, we found an association between the proportions of agriculture in the HU8 and freshwater cluster groups. Stream-rich clusters were associated with higher proportions of mean agriculture compared to lake- and wetland-rich areas. Agricultural activity can directly modify the freshwater landscape by diverting or extracting water, creating impoundments, and draining wetlands (Smith *et al.*, 2002; Wright & Wimberly, 2013) and may preferentially remove isolated systems (Cohen *et al.*, 2016; Rains *et al.*, 2016). However, agriculture may be correlated with other variables such as soil composition that may influence freshwater composition and connectivity attributes.

These variables may be indicators of the diverse drivers that shape freshwater composition and connectivity characteristics across the landscape. At these macroscales, past geological activity (i.e., glacial and fluvial processes) may be key drivers that affect the presence of freshwater systems; and hydrologic, geologic, and human land use activity may be more variable in how they modify freshwater characteristics across regions. It should be noted that the objectives of this exploratory analysis were to identify potential geographic variables that may be associated with freshwater abundance and connectivity characteristics. The associations do not imply causative relationships but rather highlight potential variables to examine in greater depth, at finer spatial scales in relation to the freshwater landscape.

Limitations in freshwater metrics and future directions

Our freshwater connectivity metrics are simplified representations of the spatial configuration of freshwater systems in the landscape and thus do not capture all aspects of freshwater connectivity. In particular, data are lacking on groundwater, reservoirs, and dams, the metrics are only as good as the resolution and accuracy of the original data layers, and our metrics do not capture temporal changes in freshwater connectivity. Future work should expand on these limitations to provide a more complete view of the freshwater landscape.

Surface hydrologic connectivity is a spatial characteristic but the magnitude and presence of these connections are dynamic through time based on seasonal changes in climate and hydrology. This is especially true for isolated wetlands and intermittent and ephemeral streams. These systems lack persistent surface water connections producing conditions that support unique biogeochemical and biological functions (Larned *et al.*, 2010; Datry *et al.*, 2014; Cohen *et al.*, 2016; Rains *et al.*, 2016). However these systems are being modified and lost at high rates due to land use activities and climate change, and thus there is a need to assess what is currently present (Larned *et al.*, 2010; Rains *et al.*, 2016). We lacked temporal data to be able to incorporate this temporal dimension in our connectivity metrics, and thus our metrics capture more permanent freshwater connectivity patterns based on static spatial data layers. However, this presents another research frontier to integrate temporary, dynamic water bodies into the freshwater landscape.

Despite these limitations, our freshwater connectivity metrics represent the complex surface hydrologic network that links lakes, wetlands, and streams and moderates the flow of water, materials, nutrients, and organisms across the landscape. In addition, these metrics are relatively easy to calculate using widely available geospatial data and can be applied at spatial scales aligned with disturbance assessment and management. This study is one of the first attempts at measuring freshwater connectivity at broad scales and we expect our ability to do so will improve in the future as the underlying data improve; as we gather additional data that are currently lacking such as on smaller water bodies, groundwater, and dams; and as new metrics for abundance and connectivity are developed.

Policy and management implications

The patterns and distributions of the integrated freshwater landscape can inform empirical and applied science alike. While it is common for freshwater systems to be studied in isolation within disciplinary boundaries, it is becoming widely recognized that systems need to be studied together. Relationships among lakes, wetlands, and streams and their hydrologic connections (or lack thereof) to one another significantly influence critical processes such as water, nutrient, and carbon fluxes among ecosystems (e.g., Quinlan *et al.*, 2003; Cardille *et al.*, 2007; Acuña & Tockner, 2010; Lottig *et al.*, 2011), nutrient and carbon processing (Weller *et al.*, 1996; Kling *et al.*, 2000; Strayer *et al.*, 2003), and biological composition and population dynamics (Crump *et al.*, 2007; Bouvier *et al.*, 2009; Nelson *et al.*, 2009). These various processes can influence freshwater ecosystem integrity and functions, and therefore integrated freshwater connectivity has important implications for policy and management.

Because freshwater connectivity characteristics play such an integral role in physical, chemical, and biological integrity of freshwater ecosystems and because they are vulnerable to disturbances, freshwater connectivity has been proposed as key considerations to management and conservation actions (Pringle, 2001, 2003). Freshwater connectivity characteristics, which can be conceptualized as a gradient ranging from isolated to highly connected systems, can sustain hydrologic and biogeochemical conditions and support metapopulation dynamics. But, freshwater connections may also impair ecosystems by transporting nutrients, contaminants, and spreading non-native species (Pringle, 2001; Rahel, 2007), which can have broad-scale regional consequences to downstream receiving waterbodies (Freeman *et al.*, 2007). In addition, because freshwater systems that are geographically isolated from surface waters are connected to hydrologic networks through subsurface or overland flow, they can provide many of the ecosystem services that are associated with connected systems (Cohen *et al.*, 2016; Rains *et al.*, 2016). But much still remains unknown about the role of freshwater connectivity at macroscales.

Broad-scale disturbances such as land use conversion and climate change threaten the integrity and function of the freshwater landscape by altering freshwater connectivity. Urbanization and agricultural land use physically alter the size, shape, and connectivity characteristics of freshwater systems through water extraction and diversion, stream channelization, impoundment and burial of headwater bodies, wetland drainage, and altered flow regimes through dam and reservoir construction (Zedler & Kercher, 2005; Freeman *et al.,* 2007; Vörösmarty *et al.*, 2010; Carpenter *et al.*, 2011; Steele & Heffernan, 2014; Van Meter & Basu, 2015). In addition, changes in temperature and precipitation patterns associated with climate change are likely to impact freshwater systems and connectivity characteristics. However, it is difficult to assess the impact of these disturbances and prescribe appropriate management actions without recognizing the distribution of freshwater connectivity at broad-scales and the importance of freshwater connectivity or lack thereof in providing ecosystem services.

Freshwater connectivity characteristics have been used as criteria to determine which waterbodies are given protection under the U.S. Clean Water Act (Leibowitz *et al.*, 2008). Currently, protection extends to navigable waters and waterbodies with a 'significant nexus' to other navigable waters such that intermittent stream flows or geographically isolated wetlands are subject to be excluded. However, a recent report by the U.S. Environmental Protection Agency (EPA) investigated how geographically isolated and non-perennial systems my affect ecological integrity of downstream waters and concluded that more research is needed to clarify the specific role these systems may play (US EPA, 2015). To achieve these goals, there is a need to characterize the distribution of freshwater systems and their connectivity characteristics at macroscales as we have done here. A macroscale perspective can further our understanding of the importance of isolated and connected freshwater systems by examining these systems within the various landscape contexts in which they operate and to identify and target systems that need protection. In addition, macroscale patterns of the freshwater landscape are aligned with the scales at which some of the leading disturbances to freshwater systems operate such as land use conversion and climate change and may be better aligned with the scales at which management is performed.

APPENDIX

Chapter 2: Appendix Tables & Figures

Table B1. Summary statistics of hydrogeomorphic, climate, and land use variables within the HU8 scale (n = 445). Median, standard deviation (st.dev.), and range of metric values are presented below.

Variable	Metric	Median (st.dev.)	Range
Hydrogeomorphic			
	HU8 area (ha)	288038 (176639)	3612 - 1245707
	Mean slope	2.86 (2.68)	0.32 – 13.96
	Mean baseflow (%)	47.18 (14.79)	9.82 – 86.36
	Mean runoff (in/yr)	12.02 (7.09)	1.73 – 30.11
	% Alluvial	17.59 (7.41)	0 – 88.98
	% Decomposition residuum	0 (12.14)	0 – 75.67
	% Eolian silt	0 (3.13)	0 – 27.61
	% Glacial fluvial outwash	1.52 (11.58)	0 – 61.77
	% Lacustrine	0 (9.20)	0 – 57.74
	% Lacustrine clay	0 (12.02)	0 – 91.68
	% Peat marsh	0 (6.18)	0 – 67.44
	% Soliflucation	0 (9.93)	0 – 100
	% Till clay	0 (18.72)	0 – 99.96
	% Till loam	41.73 (32.86)	0 – 100
	% Till sand	0 (24.79)	0 – 97.88
Climate			
	Precipitation (30 yr. mean) (mm)	977.30 (170.51)	566.20 – 1376.60
Land use/ Land cover			
	% Agriculture	33.20 (29.80)	0 – 89.30
	% Urban	6.76 (11.41)	0.02 – 82.99
	% Pasture	7.40 (9.95)	0 – 48.83

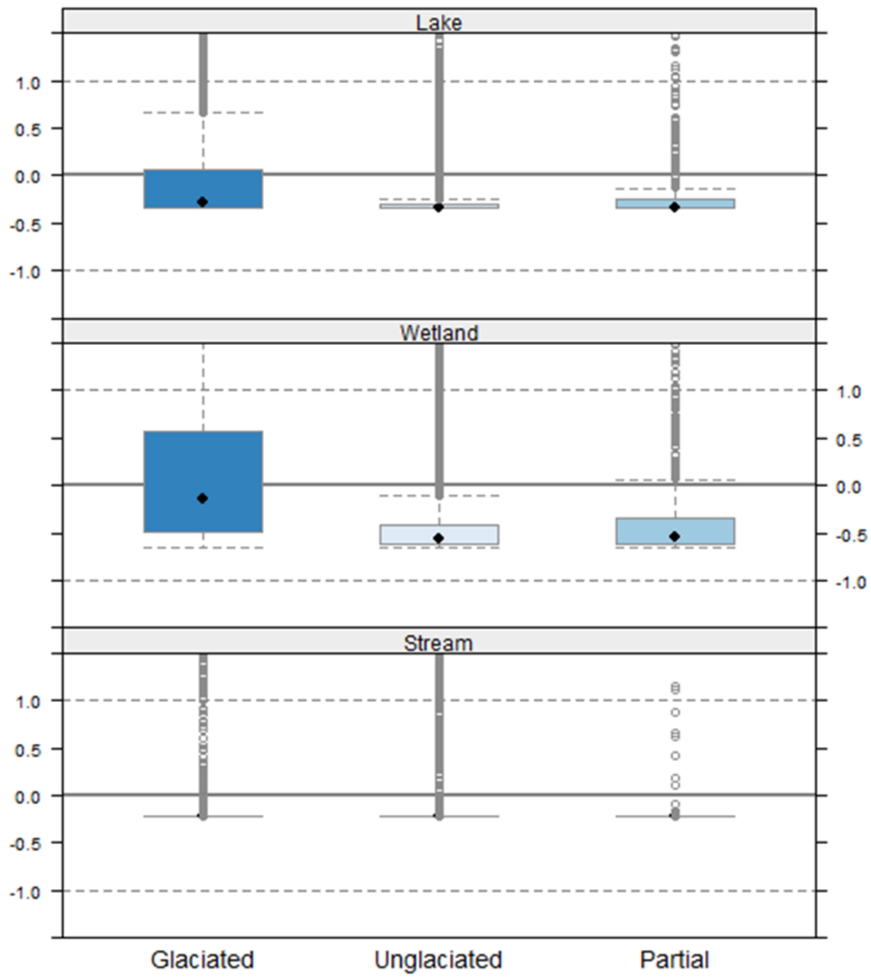

Figure B2. Distribution of freshwater system types by glaciation regime. Lake, wetland, and stream abundance proportions in the HU12 were standardized by subtracting the mean and dividing by the standard deviation. HU12 spatial units were grouped into three glaciation regimes based on the Wisconsin Glacial Stage. The solid gray line represents the mean and the dashed lines represent ±1 standard deviation above or below the mean.

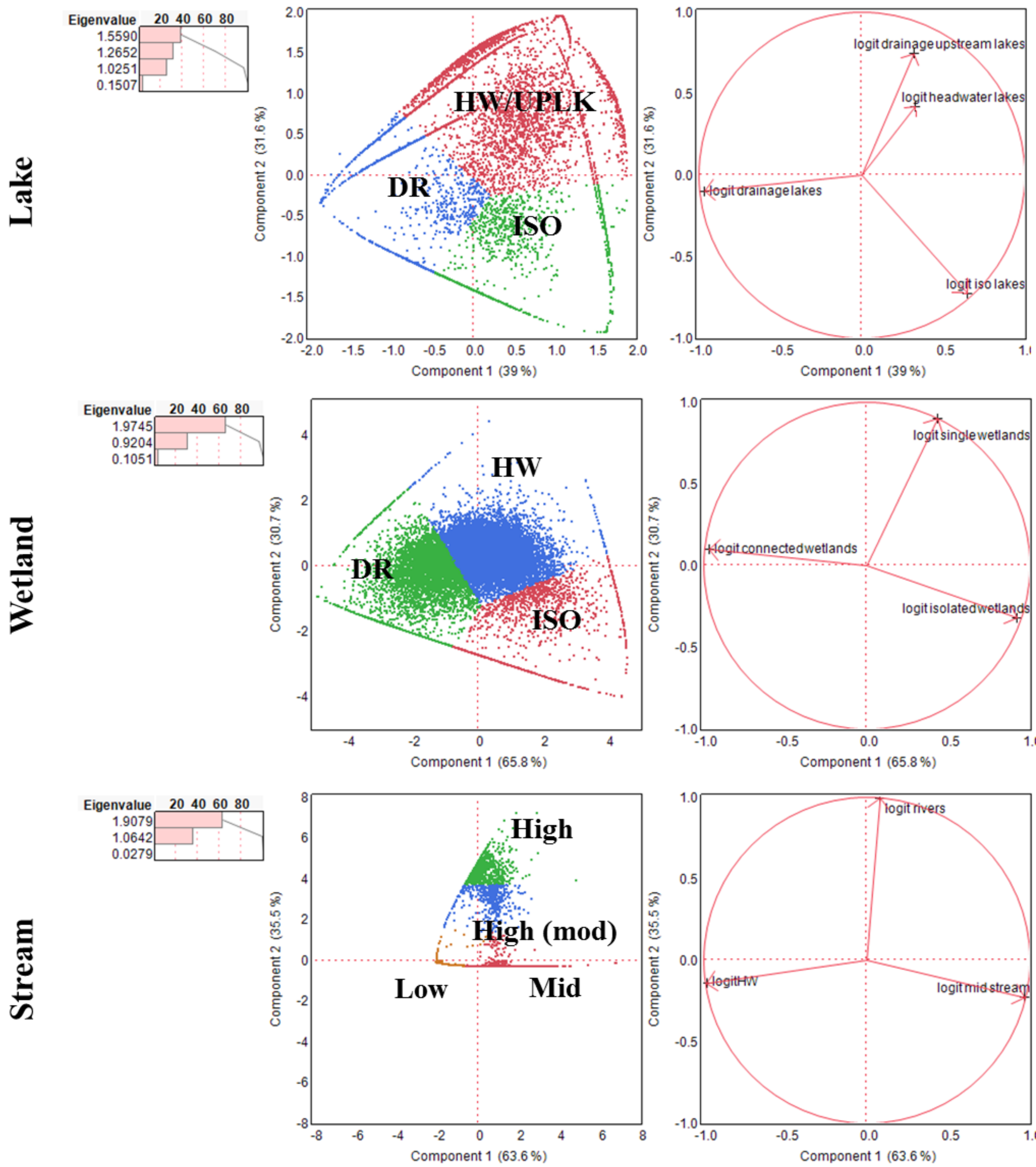

Figure B3. Principal Component Analysis (PCA) and K-Means Cluster Analysis for freshwater connectivity metrics at HU12 scale. PCA and k-means cluster analyses were performed separately for lake, wetland, and stream connectivity metrics. Cluster groups were interpreted by the dominate connectivity metric that characterized the PCA score. ISO = Isolated; HW = Headwater; DR = Drainage; UPLK = Drainage-UPLK ; Low = Low-order; Mid = Mid-order; High = High-order.

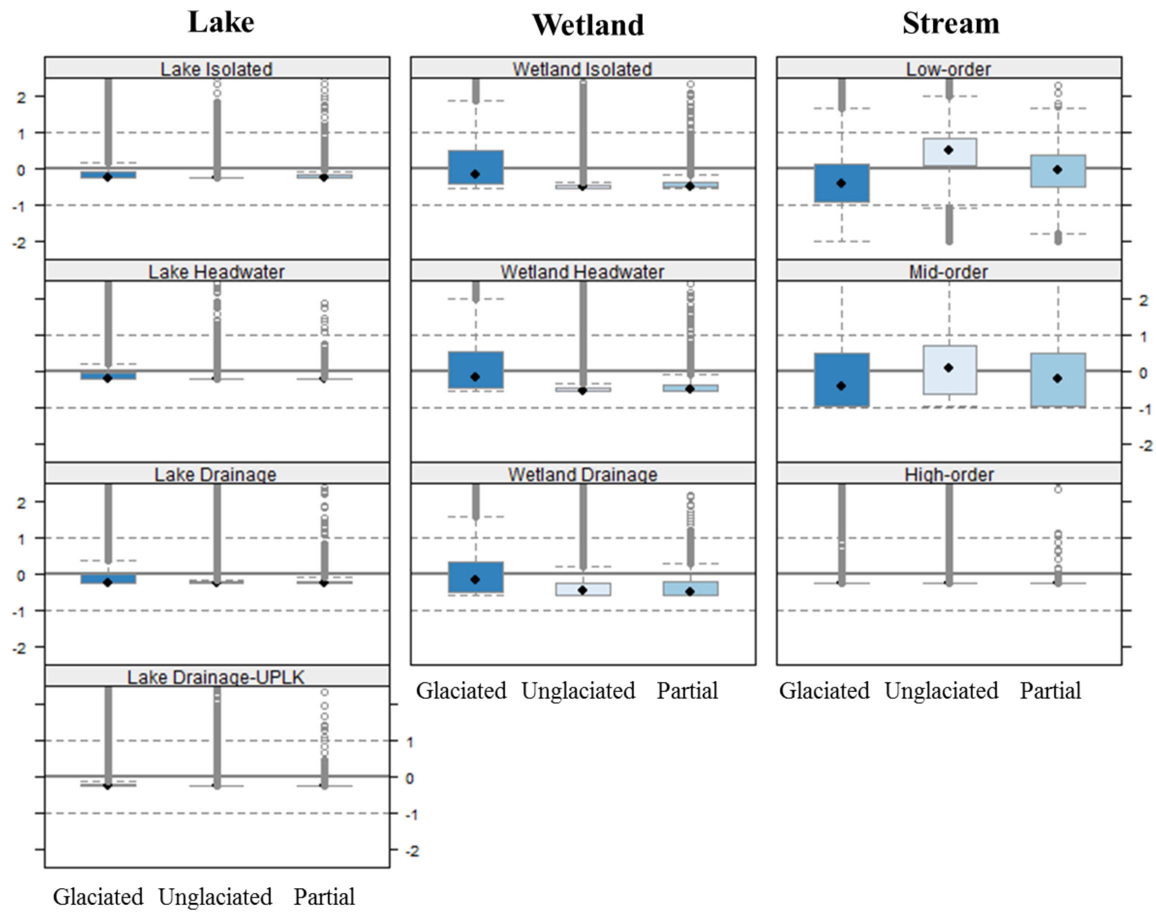

Figure B4. Distribution of freshwater connectivity types by glaciation regime. Lake, wetland, and stream connectivity type absolute proportion area in the HU12 spatial unit were standardized by subtracting the mean and dividing by the standard deviation. HU12 spatial units were grouped into three glaciation regimes based on the Wisconsin Glacial Stage: Glaciated, Unglaciated, and Partial. The solid gray line represents the mean and the dashed lines represent ±1 standard deviation above or below the mean.

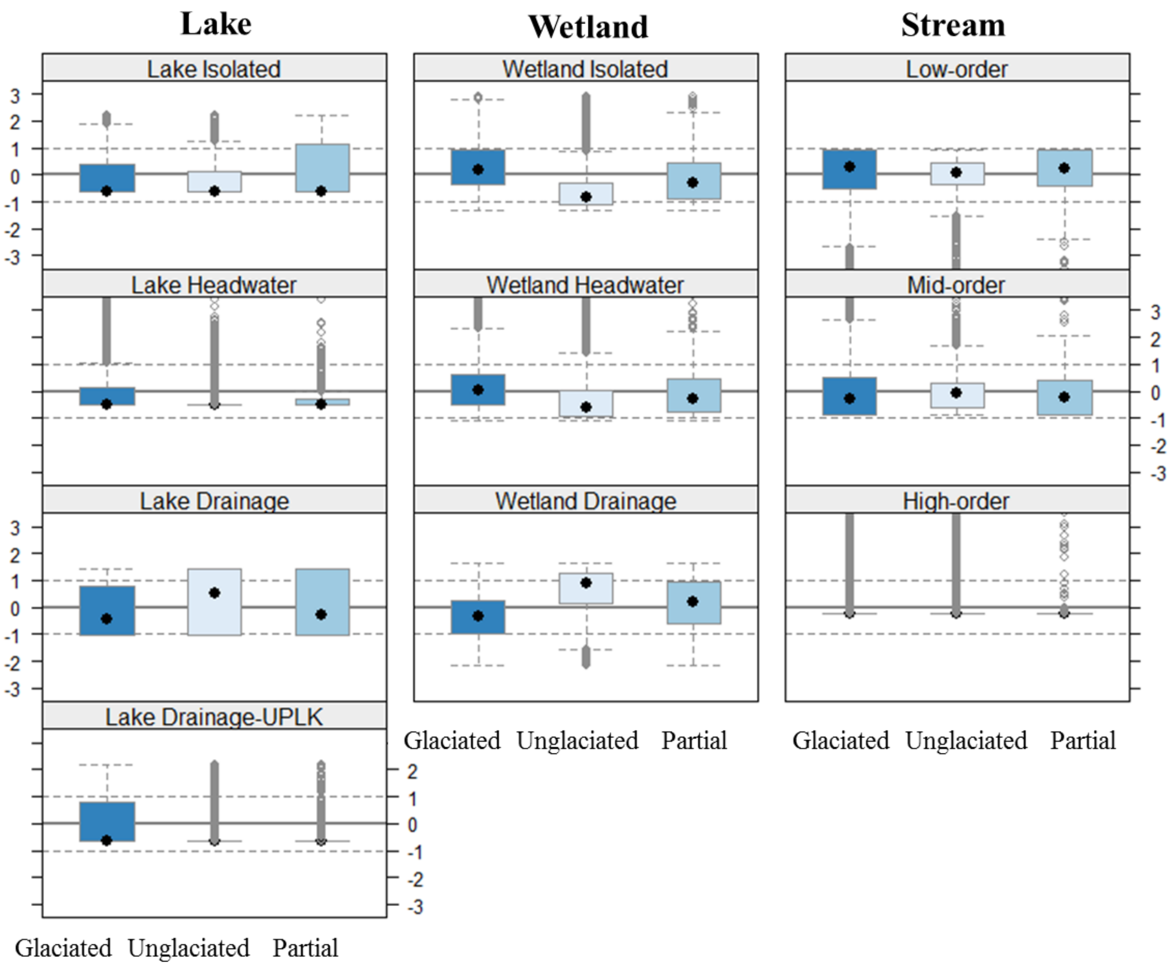

Figure B5. Distribution of relative proportion freshwater connectivity types by glaciation regime. Lake, wetland, and stream connectivity type relative proportion area in the HU12 spatial unit were standardized by subtracting the mean and dividing by the standard deviation. HU12 spatial units were grouped into three glaciation regimes based on the Wisconsin Glacial Stage: Glaciated, Unglaciated, and Partial. The solid gray line represents the mean and the dashed lines represent ±1 standard deviation above or below the mean.

Table B6. Summary statistics of freshwater metrics. Lake, wetland, and stream abundance and connectivity metrics were quantified at two Hydrologic unit (HU) spatial scales: HU12 (n = 18876 and HU8 (n = 455). Lake and wetland connectivity metrics were calculated as the relative proportion area out of the total lake or wetland area within the spatial unit. Stream connectivity was calculated as the relative density of stream connectivity type out of the total stream length within the spatial unit. Standard deviation = st. dev.

Freshwater metric type	Metric	Median (st.dev.)	Range
HU12 scale			
Lake abundance	Lake proportion	0.001 (0.05)	0 – 0.99
Lake size	Lake size (ha)	18 (566521)	0 – 8213881
Lake-stream connectivity	Isolated $_{lake}$	0 (0.35)	0 – 1.00
	Headwater $_{lake}$	0 (0.26)	0 – 1.00
	Drainage $_{lake}$	0.29 (0.40)	0 – 1.00
	Drainage-UPLK $_{lake}$	0 (0.35)	0 – 1.00
Lake-stream connectivity size	Isolated $_{lake}$ (ha)	0.103 (0.29)	0 – 8.80
	Headwater $_{lake}$ (ha)	0.13 (0.95)	0 – 36.78
	Drainage $_{lake}$ (ha)	0.17 (1.84)	0 – 77.30
	Drainage-UPLK $_{lake}$ (ha)	0.46 (5.63)	0 – 99.66
Wetland abundance	Wetland proportion	0.03 (0.11)	0 – 0.99
Wetland size	Wetland size (ha)	1.79 (10.98)	0 – 425
Wetland-stream connectivity	Isolated $_{wetland}$	0.26 (0.23)	0 – 1.00
	Headwater $_{wetland}$	0.09 (0.10)	0 – 1.00
	Drainage $_{wetland}$	0.59 (0.26)	0 – 1.00
Wetland-stream connectivity size	Isolated $_{wetland}$ (ha)	0.01 (0.05)	0 – 4.99
	Headwater $_{wetland}$ (ha)	0.02 (0.09)	0 – 4.35
	Drainage $_{wetland}$ (ha)	0.04 (0.23)	0 – 9.16
Stream abundance	Stream density (m/ha)	11.19 (5.62)	0 – 70.27
Stream connectivity	Low-order	0.90 (0.13)	0 – 1.00
	Mid-order	0.09 (0.12)	0 – 1.00
	High-order	0 (0.05)	0 – 0.99
HU8 Scale			
Lake abundance	Lake proportion	0.01 (0.04)	0 – 0.41
Lake size	Lake size (ha)	48.40 (80289)	0 - 1173449
Lake-stream connectivity	Isolated $_{lake}$	0.08 (0.18)	0 – 1.00
	Headwater $_{lake}$	0.05 (0.11)	0 – 0.95
	Drainage $_{lake}$	0.30 (0.26)	0 – 1.00
	Drainage-UPLK $_{lake}$	0.41 (0.31)	0 – 1.00

Table B6 (cont'd).

Wetland abundance	Wetland proportion	0.05 (0.11)	0 – 0.95
Wetland size	Wetland size (ha)	2.15 (4.24)	0 – 44.35
Wetland-stream connectivity	Isolated $_{wetland}$	0.27 (0.14)	0.02 – 0.68
	Headwater $_{wetland}$	0.11 (0.05)	0 – 0.27
	Drainage $_{wetland}$	0.59 (0.17)	0.22 – 0.97
Stream abundance	Stream density (m/ha)	11.21 (4.66)	1.95 – 45.12
Stream connectivity	Low-order $_{stream}$	0.87 (0.04)	0.71 – 1.00
	Mid-order $_{stream}$	0.12 (0.04)	0 – 0.29
	High-order $_{stream}$	0.001 (0.02)	0 – 0.13

0

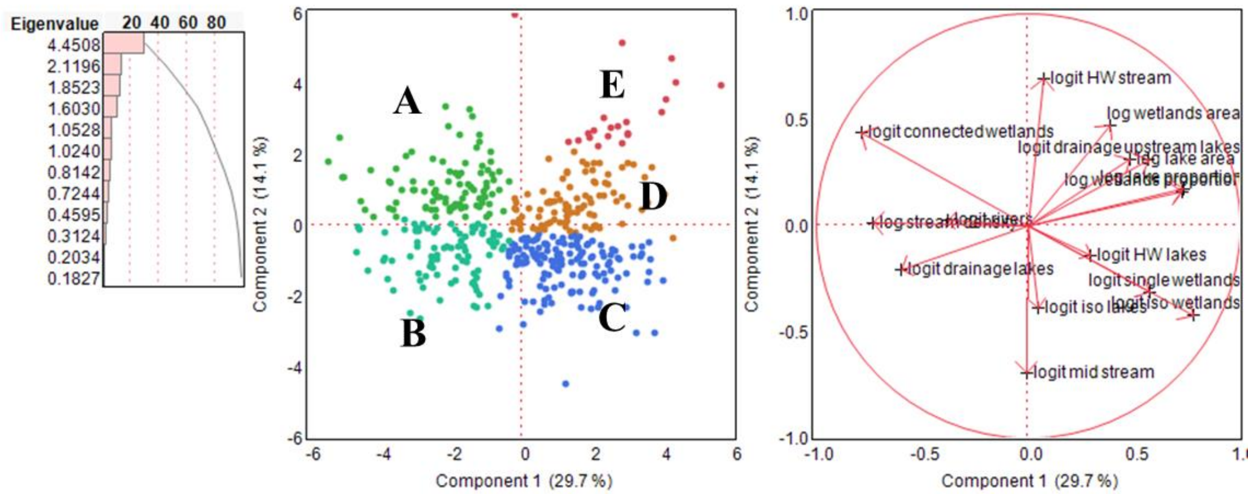

Figure B7. Principal Component Analysis (PCA) and K-Means Cluster Analysis for integrated freshwater abundance and connectivity metrics at the HU8 scale. PCA was performed on lake, wetland, and stream abundance and connectivity metrics quantified at the HU8 scale. Two PC axes were retained that explained ~43% of total variance. Five cluster groups were retained in the k-means cluster analysis (See Table 4 for descriptions of cluster groups).

Table B8. Error matrix of top performing random forest algorithm. Columns represent predicted integrated freshwater connectivity cluster groups and rows represent the instances of the actual cluster groups in the dataset. The classification error rate represents the number of instances when clusters were assigned the wrong cluster membership.

	Int_A	Int_B	Int_C	Int_D	Int_E	Total actual instances	Classification error rate
Int_A	67	18	8	1	0	94	0.29
Int_B	26	41	16	5	0	88	0.53
Int_C	0	9	118	20	2	149	0.21
Int_D	2	2	30	56	3	93	0.40
Int_E	0	0	4	14	5	23	0.78
Total predicted instances	95	70	176	96	10	–	–

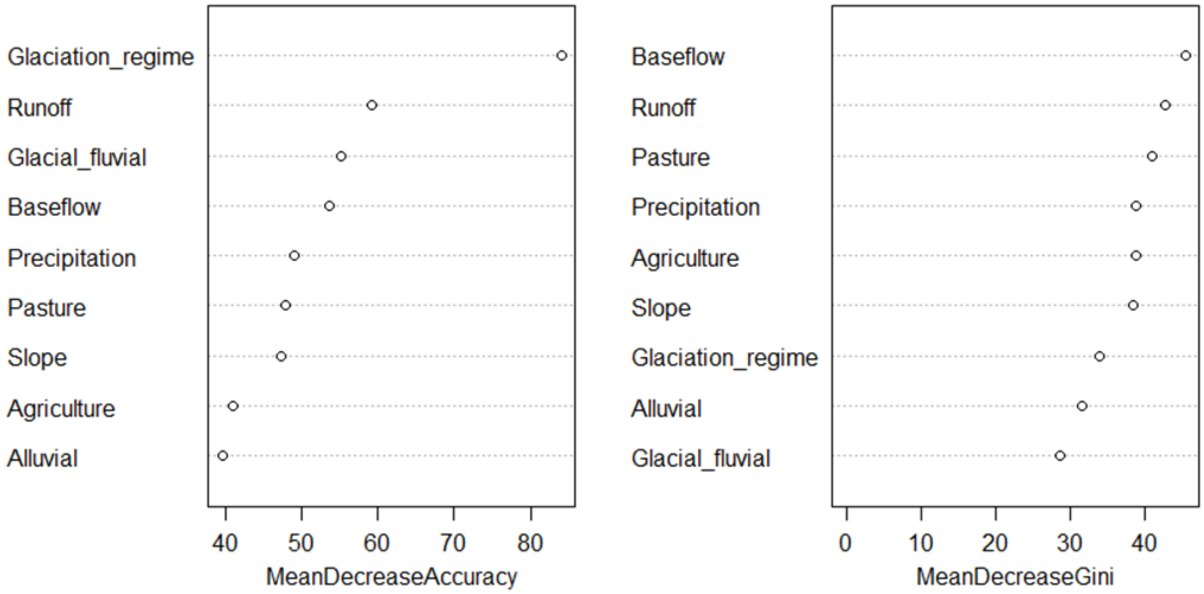

Figure B9. Random forest variable importance. The top performing random forest to predict integrated freshwater abundance and connectivity clusters assigned at the HU8 scale was selected based on the hold out prediction error rate. Hydrogeomorphic, climatic, and land use variables were included in the model and ranked based on importance in predicting cluster group assignment. The top performing random forest had a prediction error rate of 36.24%.

LITERATURE CITED

LITERATURE CITED

Acuña, V. & Tockner, K. (2010) The effects of alterations in temperature and flow regime on organic carbon dynamics in Mediterranean river networks. *Global Change Biology*, **16**, 2638–2650.

Aselmann, I. & Crutzen, P.J. (1989) Global distribution of natural freshwater wetlands and rice paddies, their net primary productivity, seasonality and possible methane emissions. *Journal of Atmospheric Chemistry*, **8**, 307–358.

Aufdenkampe, A.K., Mayorga, E., Raymond, P.A., Melack, J.M., Doney, S.C., Alin, S.R., Aalto, R.E. & Yoo, K. (2011) Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Frontiers in Ecology and the Environment*, **9**, 53–60.

Bouvier, L.D., Cottenie, K. & Doka, S.E. (2009) Aquatic connectivity and fish metacommunities in wetlands of the lower Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 933–948.

Butman, D. & Raymond, P.A. (2011) Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geoscience*, **4**, 839–842.

Butman, D., Stackpoole, S., Stets, E., McDonald, C.P., Clow, D.W. & Striegl, R.G. (2016) Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proceedings of the National Academy of Sciences*, **113**, 58–63.

Cardille, J.A., Carpenter, S.R., Coe, M.T., Foley, J.A., Hanson, P.C., Turner, M.G. & Vano, J.A. (2007) Carbon and water cycling in lake-rich landscapes: Landscape connections, lake hydrology, and biogeochemistry. *Journal of Geophysical Research: Biogeosciences*, **112**, G02031.

Carpenter, S.R., Stanley, E.H. & Zanden, M.J.V. (2011) State of the World's Freshwater Ecosystems: Physical, Chemical, and Biological Changes. *Annual Review of Environment and Resources*, **36**, 75–99.

Cohen, M.J., Creed, I.F., Alexander, L., Basu, N.B., Calhoun, A.J.K., Craft, C., D'Amico, E., DeKeyser, E., Fowler, L., Golden, H.E., Jawitz, J.W., Kalla, P., Kirkman, L.K., Lane, C.R., Lang, M., Leibowitz, S.G., Lewis, D.B., Marton, J., McLaughlin, D.L., Mushet, D.M., Raanan-Kiperwas, H., Rains, M.C., Smith, L. & Walls, S.C. (2016) Do geographically isolated wetlands influence landscape functions? *Proceedings of the National Academy of Sciences*, 201512650.

Crump, B.C., Adams, H.E., Hobbie, J.E. & Kling, G.W. (2007) Biogeography of Bacterioplankton in Lakes and Streams of an Arctic Tundra Catchment. *Ecology*, **88**, 1365–1378.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J. (2007) Random Forests for Classification in Ecology. *Ecology*, **88**, 2783–2792.

Datry, T., Larned, S.T. & Tockner, K. (2014) Intermittent Rivers: A Challenge for Freshwater Ecology. *BioScience*, bit027.

Detenbeck, N.E., Johnston, C.A. & Niemi, G.J. (1993) Wetland effects on lake water quality in the Minneapolis/St. Paul metropolitan area. *Landscape Ecology*, **8**, 39–61.

Devito, K.J., Creed, I.F., Rothwell, R.L. & Prepas, E.E. (2000) Landscape controls on phosphorus loading to boreal lakes: implications for the potential impacts of forest harvesting. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 1977–1984.

Downing, J.A., Cole, J.J., Duarte, C.A., Middelburg, J.J., Melack, J.M., Prairie, Y.T., Kortelainen, P., Striegl, R.G., McDowell, W.H. & Tranvik, L.J. (2012) Global abundance and size distribution of streams and rivers. *Inland Waters*, **2**, 229–236.

Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M. & Middelburg, J.J. (2006) The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology and Oceanography*, **51**, 2388–2397.

Fisher, S.G., Sponseller, R.A. & Heffernan, J.B. (2004) Horizons in stream biogeochemistry: flowpaths to progress. *Ecology*, **85**, 2369–2379.

Freeman, M.C., Pringle, C.M. & Jackson, C.R. (2007) Hydrologic Connectivity and the Contribution of Stream Headwaters to Ecological Integrity at Regional Scales1. *JAWRA Journal of the American Water Resources Association*, **43**, 5–14.

Hornby, D. D., 2010. RivEX software (version): Available from http://www.rivex.co.uk.

Hughes, R.M., Kaufmann, P.R. & Weber, M.H. (2011) National and regional comparisons between Strahler order and stream size. *Journal of the North American Benthological Society*, **30**, 103–121.

Humborg, C., Mörth, C.-M., Sundbom, M., Borg, H., Blenckner, T., Giesler, R. & Ittekkot, V. (2010) CO2 supersaturation along the aquatic conduit in Swedish watersheds as constrained by terrestrial respiration, aquatic respiration and weathering. *Global Change Biology*, **16**, 1966–1978.

Johnson, L., Richards, C., Host, G. & Arthur, J. (1997) Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology*, **37**, 193–208.

Jones, J.A. (2011) Hydrologic responses to climate change: considering geographic context and alternative hypotheses. *Hydrological Processes*, **25**, 1996–2000.

Jones, J.A., Creed, I.F., Hatcher, K.L., Warren, R.J., Adams, M.B., Benson, M.H., Boose, E., Brown, W.A., Campbell, J.L., Covich, A., Clow, D.W., Dahm, C.N., Elder, K., Ford, C.R., Grimm, N.B., Henshaw, D.L., Larson, K.L., Miles, E.S., Miles, K.M., Sebestyen, S.D., Spargo, A.T., Stone, A.B., Vose, J.M. & Williams, M.W. (2012) Ecosystem Processes and Human Influences Regulate Streamflow Response to Climate Change at Long-Term Ecological Research Sites. *BioScience*, **62**, 390–404.

Jones, N.E. (2010) Incorporating lakes within the river discontinuum: longitudinal changes in ecological characteristics in stream–lake networks. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 1350–1362.

Kling, G.W., Kipphut, G.W., Miller, M.M. & O'Brien, W.J. (2000) Integration of lakes and streams in a landscape perspective: the importance of material processing on spatial patterns and temporal coherence. *Freshwater Biology*, **43**, 477–497.

Kratz, T., Webster, K., Bowser, C., Maguson, J. & Benson, B. (1997) The influence of landscape position on lakes in northern Wisconsin. *Freshwater Biology*, **37**, 209–217.

Larned, S.T., Datry, T., Arscott, D.B. & Tockner, K. (2010) Emerging concepts in temporary-river ecology. *Freshwater Biology*, **55**, 717–738.

Lehner, B. & Döll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, **296**, 1–22.

Leibowitz, S.G., Wigington, P.J., Rains, M.C. & Downing, D.M. (2008) Non-navigable streams and adjacent wetlands: addressing science needs following the Supreme Court's Rapanos decision. *Frontiers in Ecology and the Environment*, **6**, 364–371.

Lohse, K.A., Brooks, P.D., McIntosh, J.C., Meixner, T. & Huxman, T.E. (2009) Interactions Between Biogeochemistry and Hydrologic Systems. *Annual Review of Environment and Resources*, **34**, 65–96.

Lottig, N.R., Stanley, E.H., Hanson, P.C. & Kratz, T.K. (2011) Comparison of regional stream and lake chemistry: Differences, similarities, and potential drivers. *Limnology and Oceanography*, **56**, 1551–1562.

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

McDonald, C.P., Rover, J.A., Stets, E.G. & Striegl, R.G. (2012) The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. *Limnology and Oceanography*, **57**, 597–606.

Meybeck, M. (1995) *Global Distribution of Lakes*. *Physics and Chemistry of Lakes* (ed. by P.A. Lerman), P.D.D.M. Imboden), and P.J.R. Gat), pp. 1–35. Springer Berlin Heidelberg.

Moore, J.W. (2015) Bidirectional connectivity in rivers and implications for watershed stability and management. *Canadian Journal of Fisheries and Aquatic Sciences*, **72**, 785–795.

Müller, R.A., Futter, M.N., Sobek, S., Nisell, J., Bishop, K. & Weyhenmeyer, G.A. (2013) Water renewal along the aquatic continuum offsets cumulative retention by lakes: implications for the character of organic carbon in boreal lakes. *Aquatic Sciences*, **75**, 535–545.

Nadeau, T.-L. & Rains, M.C. (2007) Hydrological Connectivity Between Headwater Streams and Downstream Waters: How Science Can Inform Policy1. *JAWRA Journal of the American Water Resources Association*, **43**, 118–133.

Nelson, C.E., Sadro, S. & Melack, J.M. (2009) Contrasting the influences of stream inputs and landscape position on bacterioplankton community structure and dissolved organic matter composition in high-elevation lake chains. *Limnology and Oceanography*, **54**, 1292–1305.

Pringle, C. (2003) What is hydrologic connectivity and why is it ecologically important? *Hydrological Processes*, **17**, 2685–2689.

Pringle, C.M. (2001) Hydrologic Connectivity and the Management of Biological Reserves: A Global Perspective. *Ecological Applications*, **11**, 981–998.

Quinlan, R., Paterson, A.M., Hall, R.I., Dillon, P.J., Wilkinson, A.N., Cumming, B.F., Douglas, M.S.V. & Smol, J.P. (2003) A landscape approach to examining spatial patterns of limnological variables and long-term environmental change in a southern Canadian lake district. *Freshwater Biology*, **48**, 1676–1697.

Racchetti, E., Bartoli, M., Soana, E., Longhi, D., Christian, R.R., Pinardi, M. & Viaroli, P. (2010) Influence of hydrological connectivity of riverine wetlands on nitrogen removal via denitrification. *Biogeochemistry*, **103**, 335–354.

Rahel, F.J. (2007) Biogeographic barriers, connectivity and homogenization of freshwater faunas: it's a small world after all. *Freshwater Biology*, **52**, 696–710.

Rains, M.C., Leibowitz, S.G., Cohen, M.J., Creed, I.F., Golden, H.E., Jawitz, J.W., Kalla, P., Lane, C.R., Lang, M.W. & McLaughlin, D.L. (2016) Geographically isolated wetlands are part of the hydrological landscape. *Hydrological Processes*, **30**, 153–160.

Raymond, P.A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P. & Guth, P. (2013) Global carbon dioxide emissions from inland waters. *Nature*, **503**, 355–9.

Read, E.K., Patil, V.P., Oliver, S.K., Hetherington, A.L., Brentrup, J.A., Zwart, J.A., Winters, K.M., Corman, J.R., Nodine, E.R., Woolway, R.I., Dugan, H.A., Jaimes, A., Santoso, A.B., Hong, G.S., Winslow, L.A., Hanson, P.C. & Weathers, K.C. (2015) The importance of lake-specific characteristics for water quality across the continental United States. *Ecological Applications*, **25**, 943–955.

Riera, J.L., Magnuson, J.J., Kratz, T.K. & Webster, K.E. (2000) A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. *Freshwater Biology*, **43**, 301–318.

Roy, A.H., Dybas, A.L., Fritz, K.M. & Lubbers, H.R. (2009) Urbanization affects the extent and hydrologic permanence of headwater streams in a midwestern US metropolitan area. *Journal of the North American Benthological Society*, **28**, 911–928.

Sadro, S., Nelson, C.E. & Melack, J.M. (2011) The Influence of Landscape Position and Catchment Characteristics on Aquatic Biogeochemistry in High-Elevation Lake-Chains. *Ecosystems*, **15**, 363–386.

Seekell, D.A. & Pace, M.L. (2011) Does the Pareto distribution adequately describe the size-distribution of lakes? *Limnology and Oceanography*, **56**, 350–356.

Seekell, D.A., Pace, M.L., Tranvik, L.J. & Verpoorter, C. (2013) A fractal-based approach to lake size-distributions. *Geophysical Research Letters*, **40**, 517–521.

Smith, S.V., Renwick, W.H., Bartley, J.D. & Buddemeier, R.W. (2002) Distribution and significance of small, artificial water bodies across the United States landscape. *Science of The Total Environment*, **299**, 21–36.

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

Soranno, P.A., Cheruvelil, K.S., Wagner, T., Webster, K.E. & Bremigan, M.T. (2015b) Effects of Land Use on Lake Nutrients: The Importance of Scale, Hydrologic Connectivity, and Region. *PLOS ONE*, **10**, e0135454.

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. & Leavitt, P.R. (1999) Spatial Variation among Lakes within Landscapes: Ecological Organization along Lake Chains. *Ecosystems*, **2**, 395–410.

Steele, M.K. & Heffernan, J.B. (2014) Morphological characteristics of urban water bodies: mechanisms of change and implications for ecosystem function. *Ecological Applications*, **24**, 1070–1084.

Strahler, A.N. (1957) Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union*, **38**, 913–920.

Strayer, D.L., Beighley, R.E., Thompson, L.C., Brooks, S., Nilsson, C., Pinay, G. & Naiman, R.J. (2003) Effects of Land Cover on Stream Ecosystems: Roles of Empirical Models and Scaling Issues. *Ecosystems*, **6**, 407–423.

Tiner, R.W. (2003) Geographically isolated wetlands of the United States. *Wetlands*, **23**, 494–516.

Tiner, R.W. (1997) NWI Maps: What they tell us. *National Wetlands Newsletter*, **19**, 7–12.

U.S., EPA. (2015) Connectivity of Streams and Wetlands to Downstream Waters: A Review and Synthesis of the Scientific Evidence (Final Report).

Van Meter, K.J. & Basu, N.B. (2015) Signatures of human impact: size distributions and spatial organization of wetlands in the Prairie Pothole landscape. *Ecological Applications*, **25**, 451–465.

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**, 2014GL060641.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R. & Davies, P.M. (2010) Global threats to human water security and river biodiversity. *Nature*, **467**, 555–561.

Webster, K.E., Soranno, P.A., Baines, S.B., Kratz, T.K., Bowser, C.J., Dillon, P.J., Campbell, P., FeE. ., E.J. & Hecky, R.E. (2000) Structuring features of lake districts: landscape controls on lake chemical responses to drought. *Freshwater Biology*, **43**, 499–515.

Weller, C.M., Watzin, M.C. & Wang, D. (1996) Role of wetlands in reducing phosphorus loading to surface water in eight watersheds in the Lake Champlain Basin. *Environmental Management*, **20**, 731–739.

Wetland Mapping Standard — Federal Geographic Data Committee.

Winter, T.C. (2000) The Vulnerability of Wetlands to Climate Change: A Hydrologic Landscape Perspective1. *JAWRA Journal of the American Water Resources Association*, **36**, 305–311.

Wright, C.K. & Wimberly, M.C. (2013) Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences*, **110**, 4134–4139.

Yuan, J., Cohen, M.J., Kaplan, D.A., Acharya, S., Larsen, L.G. & Nungesser, M.K. (2015) Linking metrics of landscape pattern to hydrological process in a lotic wetland. *Landscape Ecology*, **30**, 1893–1912.

Zedler, J.B. & Kercher, S. (2005) WETLAND RESOURCES: Status, Trends, Ecosystem Services, and Restorability. *Annual Review of Environment and Resources*, **30**, 39–74.

Chapter 3: The effect of freshwater connectivity and multi-scale interactions on lake nutrients and carbon at macroscales

Abstract

Nutrients (phosphorus and nitrogen) and colored organic carbon are major lake water chemistry components that influence lake ecological function and integrity. They are highly variable across regional to continental scales and much of this macroscale variation remains unexplained. Past research has related lake nutrients and carbon to lake and landscape attributes that are associated with one of three important mechanisms: 1) nutrient and carbon sources on the landscape, 2) transport and delivery pathways, and 3) internal processing. However, most past research has studied the terrestrial components of these three mechanisms, despite the recognition that freshwaters could influence the first two processes in important ways. We expect that freshwater features and their connectivity are important to lake water chemistry because freshwater features are sites of active biogeochemical reactions; their hydrologic connections can deliver nutrients and carbon across the landscape; and because they may interact (either at local or across scales) with source and internal processing mechanisms to indirectly affect lake nutrients and carbon. We developed multi-level mixed-effects models to predict lake total phosphorus, total nitrogen, and water color using a suite of predictor variables that were linked to the above hypothesized mechanisms (source, transport, internal processing of both terrestrial and freshwater characteristics, and interactions with freshwater abundance and their connectivity) that control lake nutrients and carbon concentrations. We tested for hypothesized local-scale and cross-scale interactions between the mechanistic factors that control lake nutrients (i.e., internal vs. source, internal vs. transport, transport vs. source) by comparing candidate models with the above

relationships and evaluating them using the information criteria of the candidate models. We found that freshwater features and their connectivity affect lake water chemistry with evidence for both local and cross-scale interactions, but these relationships vary depending on the water chemistry variable of interest. Our results demonstrate that there is a need to integrate freshwater features and connectivity at multiple scales to explain macroscale variation in freshwater nutrients and organic carbon.

Introduction

Phosphorus, nitrogen, and colored dissolved organic carbon are highly variable in freshwater systems across regional to continental scales (Sobek *et al.*, 2007; Taranu & Gregory-Eaves, 2008; Webster *et al.*, 2008; Seekell *et al.*, 2014; Read *et al.*, 2015). To account for spatial variation in lake water chemistry within regions, comparative studies have found that watershed geomorphology, lake depth, and landscape features are related to water chemistry (Rasmussen *et al.*, 1989; D'Arcy & Carignan, 1997; Rantakari *et al.*, 2004). These empirical studies can be used to infer underlying mechanisms influencing lake water chemistry by identifying the landscape predictors that are related to three important processes that affect lake nutrients and carbon: 1) nutrient and carbon sources in lake watersheds, 2) transport and delivery pathways, and 3) internal processing within the lake (Haygarth *et al.*, 2005; Laudon *et al.*, 2011). For example, the relationships between non-point source land use and freshwater nutrients have been extensively examined and have provided insight on how different land use activities can promote variation in nutrient composition (Carpenter *et al.*, 1998; Arbuckle & Downing, 2001). In addition, it is well established that lake depth influences hydrologic dynamics like water retention time which in turn affects internal processing of nutrients and carbon (D'Arcy & Carignan, 1997; Webster *et al.*, 2008). And lake drainage ratio, which is the ratio of watershed surface area to lake surface area, has also been shown to be an indicator or metric of water retention time (Kalff, 2002). However, the effect of transport processes that link landscape nutrient and carbon sources to recipient waters, often via transport by connected lakes, streams, and wetlands to downstream water bodies has been less studied and is poorly understood (Haygarth *et al.*, 2005; Laudon *et al.*, 2011).

Freshwater systems are active components of the landscape that process, store, and transport nutrients and carbon to downstream systems. In addition, freshwater systems and their connectivity are not randomly distributed across the landscape, but rather exhibit distinct macroscale patterns (Meybeck, 1995; Smith *et al.*, 2002; Butman & Raymond, 2011, Fergus et al. in prep) that may promote interactions within and across scales. In fact, freshwater features and their connectivity have been shown to interact with watershed characteristics to influence water chemistry in stream dissolved organic carbon (DOC) (Laudon *et al.*, 2011), lake nitrogen and phosphorus concentrations in agriculture-dominated landscapes (Fraterrigo & Downing, 2008), and nitrogen processing (Seitzinger *et al.*, 2006).

In addition, freshwater systems are influenced not only by features in the local watershed but also by regional features that can affect the composition and ecological processes in freshwater systems (Soranno *et al.*, 2009; Fergus *et al.*, 2011; Wagner *et al.*, 2011; Cheruvelil *et al.*, 2013; Filstrup *et al.*, 2014). Local watershed and regional features that drive variation in water chemistry can interact and promote ecological complexity. These cross-scale interactions (CSI), defined as processes at one spatial or temporal scale interacting with processes at another scale, can make it challenging to extrapolate relationships to new areas and make predictions (Peters *et al.*, 2007). There are few examples of CSIs in freshwater systems (but, see Fergus *et al.*, 2011; Wagner *et al.*, 2011; Filstrup *et al.*, 2014; Soranno *et al.*, 2014) and relatively little is known about what may promote these macroscale relationships. It has been proposed that connectivity is associated with CSIs, whereby intermediate-scaled transfer processes may affect how pattern-process relationships interact (Peters *et al.*, 2007, 2008). And, because freshwater connectivity affects the distribution of materials, energy, and organisms within and across spatially distributed landscape elements, it may promote CSIs in freshwater systems.

We hypothesize that freshwaters and their connectivity directly affect lake nutrients and carbon and may interact with source and internal processing mechanisms at local scales and across scales to affect lake nutrients and carbon (Fig. 10). Our research questions are: 1) How do the abundance of freshwaters in the landscape (e.g., wetlands, streams, groundwater) and their connectivity affect lake total phosphorus (TP), total nitrogen (TN), and water color at macroscales? And 2) Does the abundance of freshwater systems and their connectivity interact with hypothesized drivers controlling nutrient and carbon in lakes (i.e., landscape sources and internal processing) to affect lake TP, TN, and water color? If so, at what scale(s) do these interactions occur?

To address these questions, we developed multi-level models to predict lake TP, TN, and water color using a suite of watershed and regional features that are linked to hypothesized mechanisms (source, transport, and internal processing) controlling lake nutrients and carbon concentrations. Variables that were related to nutrient sources included agriculture and urban land uses; and the variables related to carbon sources were forest and wetland land cover. Variables related to internal processing were lake depth and drainage ratio. Variables related to transport processes were freshwater abundance measures that included wetlands, streams, and baseflow (an indicator of groundwater contribution to surface waters), and freshwater connectivity measures that included stream density and lake- and wetland-stream connectivity types, ranging from isolated systems (no stream connections) to drainage systems (stream-connected). Separate analyses were performed by lake connectivity type to look at whether lake connectivity types had different drivers of nutrients and carbon.

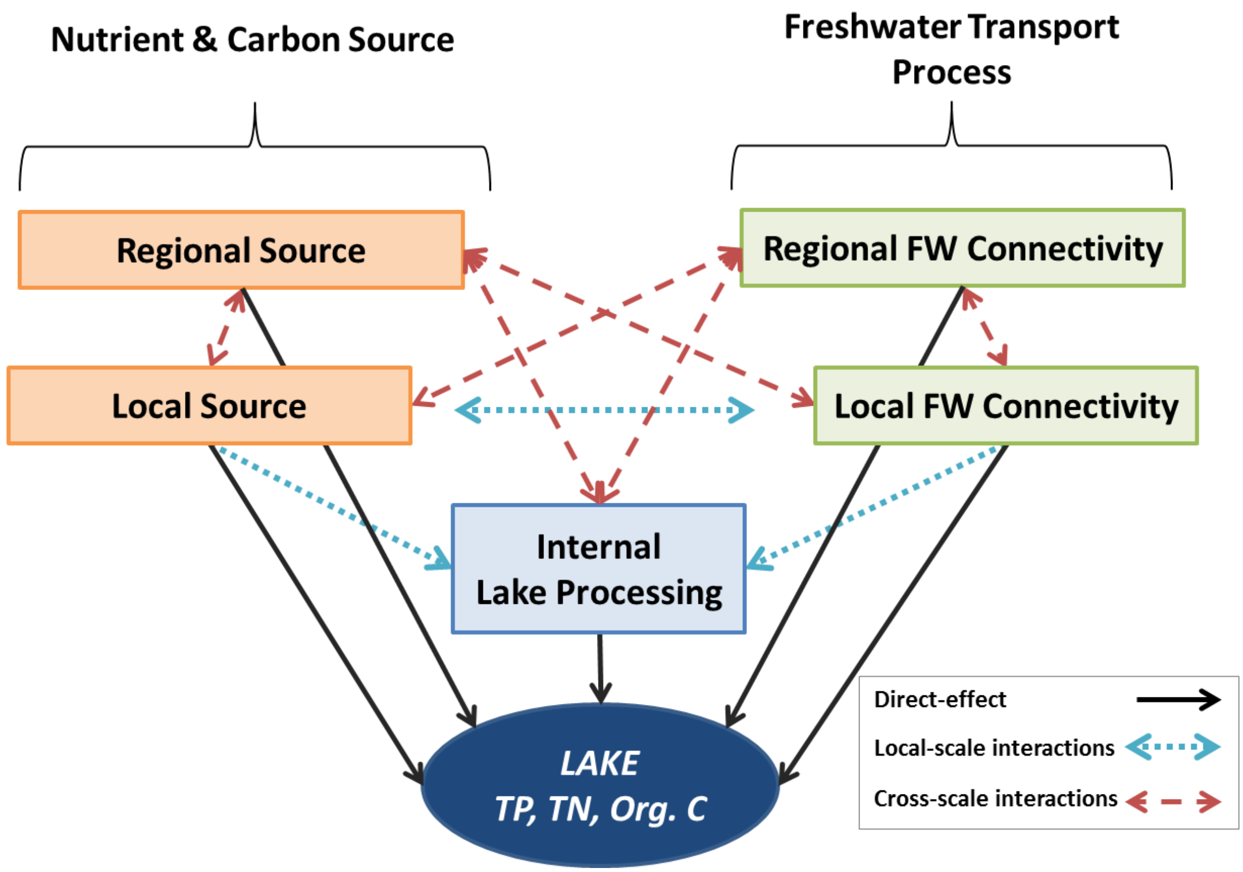

Figure 10. Conceptual framework of hypothesized multi-scaled mechanisms of lake water chemistry. Mechanistic processes are grouped into three types: Nutrient and Carbon Source; Freshwater Transport Process, and Internal Lake Processing. All three process types are hypothesized to have direct effects on lake total phosphorus (TP), total nitrogen (TN), and colored organic carbon (Org. C) as indicated by the solid dark arrow. Landscape sources and freshwater transport processes – represented by freshwater (FW) connectivity – may operate at both local and regional spatial scales; internal processing operates within the lake at local scales. Within-scale local interactions between the three processes types (dotted blue double arrows) and cross-scale interactions between local-scale and regional-scale process types (dashed red double arrows) are hypothesized to influence lake water chemistry depending on the response variable.

Methods

Lake datasets

Lake TP, TN, and water color data come from the LAGOS database (Lake multi-scaled geospatial and temporal database, Soranno *et al.*, 2015). LAGOS is a multi-thematic lake database that integrates lake water chemistry ($LAGOS_{LIMNO}$) and geospatial data ($LAGOS_{GEO}$)

across the U.S. Midwest and Northeast regions. We accessed $LAGOS_{LIMNO}$ version 1.0540.1 and $LAGOS_{GEO}$ version 1.03 for this study.

We used a subset of lakes from the LAGOS dataset, those with water chemistry and lake geomorphology data related to our research questions. We created three datasets – TP, TN, and water color to use in the analyses. Our datasets included lakes (greater than 4 ha in surface area) that had summer (15 June – 15 September) epilimentic water chemistry observations. We included lake TP, TN, and water color measurements conducted by state agencies from 1990 to 2013 following standard field and laboratory methods. For lakes that had multiple observations over time, we calculated the median annual summer water chemistry value. Lakes and their watershed were nested within HU4 hydrologic units which we treated as the region-level in our analyses. We dropped lakes missing maximum depth data and lakes missing wetland data at the watershed scale and the region scale from the analysis. The lake TP dataset included n = 4711 lake observations nested within N = 47 HU4 regions (Fig. 11a). The lake TN dataset included n = 1567 lake observations nested within N = 46 HU4 regions (Fig. 11b). The lake water color dataset included n = 1624 lake observations nested within N = 43 HU4 regions (Fig. 11c). The individual lake datasets captured a wide range in response variable values (Table 6).

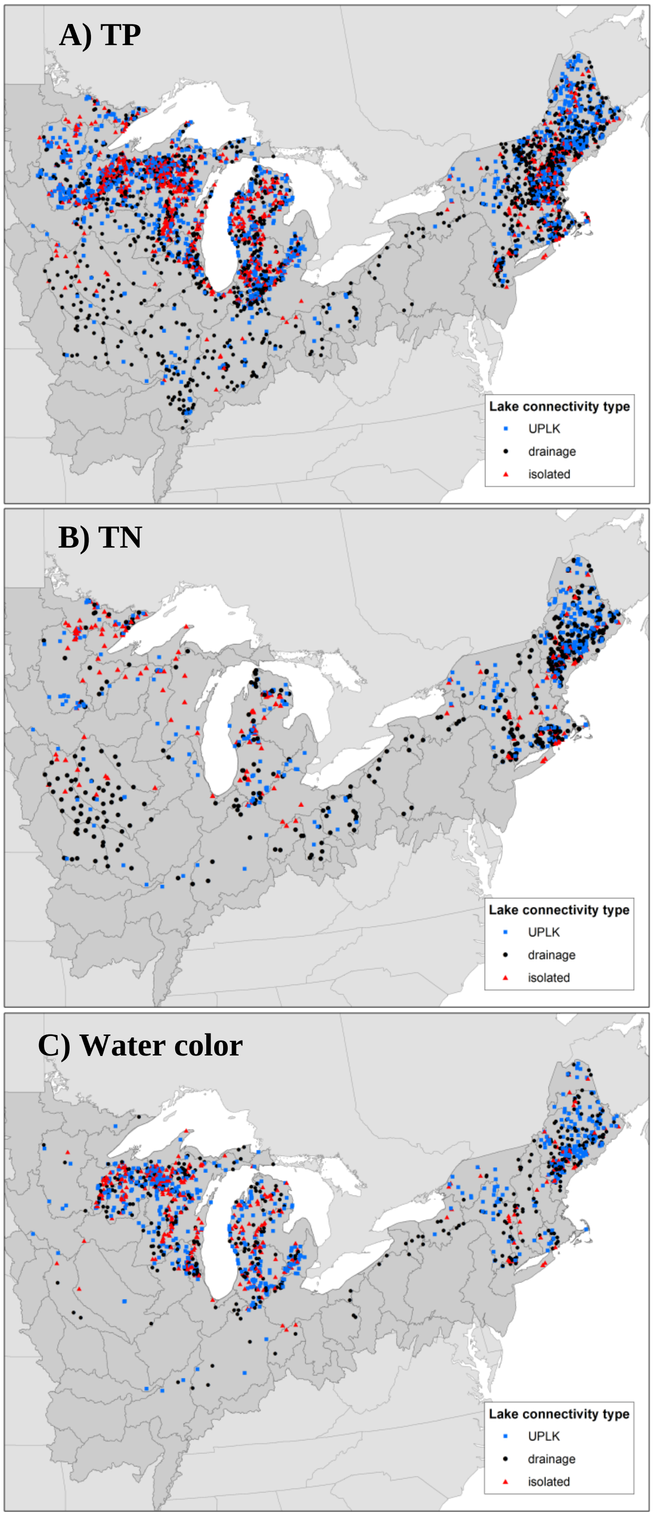

Figure 11. Study extent and locations of lakes by lake connectivity type for each dataset. A) Lakes with total phosphorus – TP (n = 4711; N = 47 HU4 regions). **B)** Lakes with total nitrogen – TN (n = 1567; N = 46 HU4 regions); **C)** Lakes with water color (n = 1624; N = 43). Isolated lakes have no inflowing streams. Drainage lakes have inflowing streams. UPLK lakes have inflowing streams and upstream lakes.

Table 6. **Summary statistics of lake total phosphorus (TP), total nitrogen (TN), and water color**. Summary statistics are presented for the all lake type dataset for TP (n =4711), TN (n=1567), and water color (n = 1624). PCU = Platinum cobalt units

Response variable	Mean	Median	Range	Standard deviation
TP μg/L	35	16	0.15 – 1476	68
TN μg/L	690	469	62 – 13090	991
Water color PCU	24	15	0 – 320	26

Lake and landscape predictor variables

To address our research questions, we related lake response variables to a suite of lake and landscape predictor variables aligned with hypothesized mechanistic processes – *internal processing*, *nutrient and carbon sources*, and *transport processes* (Table 7). These variables were quantified at two spatial scales: local (i.e., lake watershed) and regional (i.e., HU4 region). *Internal lake processing* was characterized by morphometry: maximum lake depth and drainage ratio (CA:LK). *Nutrient and carbon sources* included agriculture and urban land use for nutrients and forest and wetland land cover for carbon. *Transport processes* were represented in a couple of ways by freshwater systems and their connectivity characteristics. Wetland cover and stream density were quantified at local and regional scales for each lake. Wetlands were grouped into connectivity types – Isolated and Drainage. Isolated wetlands were not connected to stream segments, and Drainage wetlands were intersected by streams. We measured wetlands as total wetland percent and the percent of wetland surface area by connectivity type. Baseflow, an indicator of groundwater contribution, was measured at the regional scale. In addition, lakes were grouped into connectivity types based on the spatial relationships with streams and upstream lakes. Isolated lakes had no inflowing streams; Drainage lakes had inflowing streams and no upstream lakes, and Drainage-UPLK lakes had inflowing streams that were connected to upstream lakes (≥ 10 ha in size). We performed analyses using the 'all lake type' dataset as well as by lake connectivity type (results presented in the Appendix) to examine whether lake connectivity transport characteristics alter the predictors and interactions related to lake response variables.

Table 7. Lake and landscape predictor variables linked to hypothesized mechanistic processes.

Mechanistic process	Variable	Scale of measurement: Local (lake, watershed)	Scale of measurement: Regional (HU4)
Internal lake processing	Maximum lake depth (m)	X	
	CA:LK	X	
Nutrient landscape source	Agriculture %	X	X
	Urban %	X	X
Carbon landscape source	Forest %	X	X
	Total wetland %	X	X
Freshwater transport process	Total wetland %	X	X
	Isolated wetland %	X	
	Drainage wetland %	X	
	Stream density (m/ha)	X	X
	Baseflow index		X
	Lake connectivity type	–	–
	Isolated		
	Drainage		
	Drainage-UPLK		

Analyses

We developed five candidate multi-level models to examine the influence of freshwater features and local- and cross-scale interactions among hypothesized mechanistic processes on lake nutrients and carbon for each of the lake TP, TN, and water color datasets. Multi-level models explicitly incorporate hierarchical structure in the data (local and regional scale) and allow for cross-scale interactions to be modeled, and thus are well suited to the research questions. We quantified and compared the supporting evidence for each candidate model to increase inferential understanding of the influence of freshwater features and multi-scale interactions on lake water chemistry. Lake response variables were natural log transformed. Lake and landscape predictor variables were centered (i.e., subtract the grand mean from value) and standardized (i.e., divide the centered value by the standard deviation) to facilitate interpretation of the relative effects of the predictor variables and to reduce correlations between regional intercepts and slopes (Gelman & Hill, 2007). The following analytical steps were taken in the model building process.

Before developing the candidate models, we evaluated the hypothesis that lakes within the same HU4 region have similar water chemistry compared to lakes from other HU4 regions by fitting unconditional multi-level models that allowed for the mean of the response variable (log transformed) to vary by HU4 region (i.e., random intercept model) and did not include any predictor variables. The unconditional model partitions the total variation in the response variable into two variance components – σ^2, within-region variation and τ_{00} among-region variation. The intraclass correlation coefficient (ρ) uses these variance estimates to calculate the proportion of the total variance that is among regions and indicates the degree to which individual lake measurements within a HU4 region are correlated to one another:

$$\rho = \tau_{00} / (\tau_{00} + \sigma^2)$$

The first set of candidate models (one for TP, TN, and water color) included local-scale lake and watershed predictor variables that have been related to lake nutrients and carbon in the literature as well as freshwater variables – wetland percent and stream density. Variables were retained in the other candidate models if the 95% confidence interval for the model coefficients that did not overlap zero.

The second set of candidate models included local-scale interactions between lake and landscape variables linked to hypothesized mechanistic processes. We examined interactions between local source and local freshwater transport, local source and internal processing, and local freshwater transport and internal processing. Significant interactions were kept in the other candidate models.

The third set of candidate models allowed for local predictors to have variable effects by HU4 region (random slopes). We treated local source variables (agriculture – for TP and TN; forest – for water color), internal processing (lake depth), and freshwater transport (total wetland, *Isolated* wetland, and *Drainage* wetland) as separate random effects.

The fourth set of candidate models included HU4 regional-scale predictor variables that represented sources or transport processes. For lake nutrients TP and TN, percent agriculture in the HU4 was highly correlated with percent agriculture in the watershed, so we did not include regional agriculture in the models (see Appendix Table C11). Regional variables that included baseflow, forest, wetland, and stream density were related to lake response and significant variables were retained in the final candidate model.

The fifth and final set of candidate models added a cross-scale interaction between the random local-scale variable (slope) and the regional-scale variable. These cross-scale interactions were included to account for regional variation in the random slopes. We performed likelihood ratio tests to determine whether the model with CSIs were significantly different from models without CSIs.

An example of a model predicting lake TP is provided below that includes local- and regional-scale fixed effects, local-scale interactions, random local slopes, and cross-scale interactions. All models in the analyses were simplified variations of this model.

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(\text{Wetland}_{ij}\right) + \beta_{2j}(\text{Wetland}_{ij} \times \text{Agriculture}_{ij}) + \beta_{3j}\left(\text{Depth}_{ij}\right) + \beta_{4j}(\text{Depth}_{ij} \times \text{Regional baseflow}) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Regional\ baseflow)_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

$$\beta_{4j} = \gamma_{31}$$

where $r_{ij} \sim N\ (0, \sigma^2)$ and $\begin{pmatrix} u_{0j} \\ u_{3j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$

In this model, lake TP (Y_{ij} for lake *i* in region *j*) is a function of the overall intercept (γ_{00}), the main effect of regional agriculture (γ_{01}), fixed effect of local wetlands (γ_{10}), fixed effect of within-scale interactions between local wetlands and local agriculture (γ_{20}), random effect of depth (γ_{30}), and the cross-scale interaction between local depth and regional baseflow (γ_{31}). The intercept and lake depth slope are allowed to vary among HU4 regions by including the error terms u_{0j} and u_{3j}, where u_{0j} is the regional intercept error for region *j* and τ_{00} represents the among-region variability in lake TP after controlling for regional agriculture; u_{3j} is the regional error to the slope associated with region *j* and τ_{11} represents the among-region local depth effect variability; and τ_{01} is the covariance between u_{0j} and u_{3j}. The residual error (r_{ij}) is considered to be normally distributed (*N*) with a mean of zero and variance σ^2.

Models were evaluated based on Akaike information criteria values (AIC). Candidate models were ranked by the difference in AIC values from the lowest AIC model (Δ_i) where lower values indicate better model fit to the data. Multi-level model analyses were performed using the lme4 and nlme packages in R (Bates *et al.*, 2014; J. Pinheiro *et al.*, 2016). Parameters were estimated using restricted maximum likelihood.

Results

Regional variation in lake TP, TN, and water color

For all three lake response variables (TP, TN, and water color), there was significant among-region variation across HU4 regions as indicated by the intraclass correlation coefficient – ICC (Table 8). For the all lake type datasets, lake TP and TN had ICC values greater than 50%, indicating that over half the total variation in lake nutrients occur at the regional scale. In contrast, the water color ICC value was less than 9% with the majority of water color variation occurring at the lake level (i.e., within-region), indicating that there is a great deal of variation in water color within regions in this study extent. We found similar ICC values across the different lake connectivity types (Appendix Table C1). However, ICC values tended to be higher in Drainage and Drainage-UPLK lakes compared to Isolated lakes. This pattern suggests that lake-stream connectivity may promote within-region similarity among lake water chemistry variables. The candidate models to follow all included random intercepts by HU4 region to account for these regional similarities among lake response variables.

Table 8. Unconditional mixed-effects model intraclass correlation coefficient (ICC). The ICC represents the percent of the total variation in lake response variable that occurs at the HU4 (region) scale. (n = number of lakes; N = number of HU4 regions)

Variable	ICC	n	N
TP	52	4711	47
TN	54	1567	46
COLOR	9	1624	43

Top ranked models for TP, TN, and water color

The within-region variation in TP, TN, and water color was partially explained by local-scale lake and watershed variables (Tables 9 – 11). Local predictor variables that had similar effects across all three response variables included morphometric characteristics representing internal lake processing. Lake depth was negatively related to both nutrients and carbon, whereas drainage ratio was positively related to all three lake response variables. We also found that measures of freshwater presence or connectivity were important for all three response variables. In particular, local wetlands in the watershed were positively related to TP, TN, and water color, indicating that wetlands may act as a source of both nutrients and carbon to lakes.

However, not all variation occurred at the lake-level. The top ranked models for lake TP, TN, and water color included random intercepts and random slopes indicating that the response variables and the effect of local predictors on the response variables were regionally structured. The regional average for each of the nutrients (i.e., random intercept) exhibited spatial patterns that were different for TP, TN, and water color (Figs. 3, 5, 6). In addition, the type of local predictor with random slopes in the top ranked models was different among lake response variables. For TP, allowing for lake depth to vary by regions improved model fit; whereas for TN and water color allowing for local wetland effects to vary by regions improved model fit. We found additional landscape predictor variables and interactions that were different for lake TP, TN, and water color, which suggest that the multi-scaled mechanistic processes influencing macroscale patterns of lakes are different for nutrients and carbon.

Table 9. Mixed-effects models predicting total phosphorus (TP; n = 4711 lakes, N = 47 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TP (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Models: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	3.05 (2.92, 3.18)	3.04 (2.92, 3.18)	3.05 (2.92, 3.19)	3.00 (2.86, 3.13)	2.95 (2.82, 3.09)
Internal process $_{Local}$	Depth	-0.30 (-0.32, -0.28)	-0.30 (-0.32, -0.29)	-0.33 (-0.37, -0.29)	-0.33 (-0.37, -0.28)	-0.30 (-0.34, -0.27)
	CA:LK	0.12 (0.10, 0.14)	0.12 (0.10, 0.14)	0.12 (0.10, 0.14)	0.12 (0.10, 0.14)	0.12 (0.10, 0.14)
Source $_{Local}$	Agriculture	0.38 (0.35, 0.40)	0.38 (0.35, 0.40)	0.36 (0.33, 0.39)	0.36 (0.33, 0.39)	0.36 (0.33, 0.38)
	Urban	0.19 (0.17, 0.21)	0.19 (0.17, 0.21)	0.19 (0.17, 0.21)	0.19 (0.17, 0.21)	0.18 (0.16, 0.21)
FW Connectivity $_{Local}$	Stream	0.09 (0.07, 0.11)	0.08 (0.07, 0.11)	0.09 (0.07, 0.11)	0.09 (0.06, 0.11)	0.09 (0.06, 0.11)
	Wetland	0.11 (0.09, 0.13)	0.11 (0.09, 0.13)	0.11 (0.09, 0.13)	0.11 (0.09, 0.13)	0.11 (0.09, 0.13)
Source $_{Local}$ × Internal process $_{Local}$	Agriculture × Depth	–	-0.05 (-0.07, -0.03)	-0.02 (-0.04, 0.01)	–	–
Random effects						
Internal process $_{random}$	Depth $_{random}$	–	–	Depth $_{random}$	Depth $_{random}$	Depth $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.09 (-0.19, 0.01)	-0.20 (-0.30, -0.09)
Cross-scale interaction						
Internal process $_{Local}$ × FW Connectivity $_{Region}$	Depth $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	0.08 (0.05, 0.12)
Variance components	σ^2	0.35	0.35	0.34	0.34	0.34
	τ_{00}	0.2	0.19	0.21	0.18	0.16
	τ_{11}	–	–	0.01	0.02	0.01
Variation explained	Within region	39%	39%	41%	41%	41%
	Among region	–	–	–	71%	74%
	Total	54%	55%	54%	57%	58%
Model selection	AIC	8627	8608	8550	8547	8536
	Δ_i	91	72	14	11	0

Table 10. Mixed-effects models predicting total nitrogen (TN; n = 1567 lakes, N = 46 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TN (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Models: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	6.19 (6.11, 6.27)	6.20 (6.12, 6.28)	6.20 (6.12, 6.27)	6.19 (6.12, 6.27)	6.20 (6.12, 6.28)
Internal process $_{Local}$	Depth	-0.19 (-0.21, -0.16)	-0.19 (-0.21, -0.17)	-0.19 (-0.21, -0.17)	-0.19 (-0.21, -0.17)	-0.19 (-0.21, -0.17)
	CA:LK	0.06 (0.04, 0.09)	0.06 (0.04, 0.09)	0.06 (0.04, 0.08)	0.06 (0.04, 0.08)	0.06 (0.04, 0.08)
Source $_{Local}$	Agriculture	0.38 (0.35, 0.42)	0.38 (0.35, 0.41)	0.37 (0.33, 0.40)	0.36 (0.33, 0.39)	0.36 (0.33, 0.39)
	Urban	0.17 (0.14, 0.19)	0.16 (0.14, 0.19)	0.16 (0.14, 0.19)	0.16 (0.14, 0.19)	0.16 (0.14, 0.18)
FW Connectivity $_{Local}$	Stream	-0.01 (-0.04, 0.01)	–	–	–	–
	Wetland	0.10 (0.07, 0.12)	0.11 (0.08, 0.13)	0.12 (0.07, 0.17)	0.13 (0.08, 0.18)	0.13 (0.08, 0.18)
FW Connectivity $_{Local}$ × Internal process $_{Local}$	Wetland × Depth	–	0.03 (0.01, 0.05)	0.04 (0.02, 0.06)	0.04 (0.02, 0.06)	0.04 (0.02, 0.05)
Random effect						
FW Connectivity $_{random}$	Wetland $_{random}$	–	–	Wetland $_{random}$	Wetland $_{random}$	Wetland $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.06 (-0.13, 0.01)	-0.06 (-0.13, 0.01)
Cross-scale interaction						
FW Connectivity $_{Local}$ × FW Connectivity $_{Region}$	Wetland $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	-0.03 (-0.08, 0.02)
Variance components	σ^2	0.16	0.16	0.15	0.15	0.15
	τ_{00}	0.06	0.06	0.05	0.05	0.05
	τ_{11}	–	–	0.01	0.01	0.01
Variation explained	Within region	38%	38%	42%	42%	42%
	Among region	–	–	–	84%	84%
	Total	61%	61%	65%	65%	65%
Model selection	AIC	1703	1700	1664	1669	1675
	Δ_i	39	36	0	5	11

Table 11. Mixed-effects models predicting water color (Color; n = 1624 lakes, N = 43 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake Color (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

		Model				
Category	Variable	1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	2.79 (2.71, 2.86)	2.79 (2.71, 2.87)	2.80 (2.72, 2.89)	2.81 (2.73, 2.89)	2.82 (2.74, 2.89)
Internal process $_{Local}$	Depth	-0.20 (-0.24, -0.17)	-0.20 (-0.24, -0.17)	-0.20 (-0.23, -0.17)	-0.20 (-0.23, -0.17)	-0.20 (-0.23, -0.17)
	CA:LK	0.28 (0.23, 0.31)	0.28 (0.25, 0.31)	0.28 (0.25, 0.31)	0.28 (0.25, 0.31)	0.28 (0.25, 0.31)
Source $_{Local}$	Forest	0.01 (-0.04, 0.05)	–	–	–	–
FW Connectivity $_{Local}$	Stream	0.02 (-0.02, 0.05)	–	–	–	–
	Wetland	0.33 (0.30, 0.37)	0.33 (0.30, 0.38)	0.35 (0.28, 0.43)	0.36 (0.28, 0.44)	0.37 (0.30, 0.44)
FW Conn. $_{Local}$ × Int. process $_{Local}$	Wetland × Depth	–	0.02 (-0.01, 0.05)	–	–	–
Random effect						
FW Connectivity $_{random}$	Wetland $_{random}$	–	–	Wetland $_{random}$	Wetland $_{random}$	Wetland $_{random}$
Region-scale covariate						
Source $_{Region}$	Forest $_{Region}$	–	–	–	0.06 (-0.01, 0.12)	0.11 (0.04, 0.18)
Cross-scale interaction						
FW Connectivity $_{Local}$ × Source $_{Region}$	Wetland $_{Local}$ × Forest $_{Region}$	–	–	–	–	0.10 (0.03, 0.16)
Variance components	σ^2	0.39	0.39	0.38	0.38	0.38
	τ_{00}	0.05	0.05	0.05	0.04	0.04
	τ_{11}	–	–	0.03	0.03	0.03
Variation explained	Within region	39%	39%	41%	41%	41%
	Among region	–	–	–	20%	20%
	Total	36%	36%	38%	39%	39%
Model selection	AIC	3189	3196	3162	3166	3165
	Δ_i	27	34	0	4	3

Total phosphorus model

The top ranked model based on AIC values explained 58% of the variation in lake TP and included local- and regional-scale variables and cross-scale interactions (Table 9); although, other models also explained similar amounts of variation, but with substantially larger AICs. The regional-average lake TP exhibited spatial patterns with regions in the Midwest exhibiting higher than average lake TP and regions in the Northeast exhibiting lower than average TP values (Fig. 12a). These spatial patterns were somewhat contiguous and may indicate a broader extent of spatial grouping of lake TP.

The local-scale predictors followed expected relationships with lake TP. Agriculture and urban land use in the watershed, representing external nutrient sources, were positively related to lake TP. Both wetlands and stream density in the watershed were positively related to lake TP and may represent freshwater sources and pathways by which phosphorus is delivered to receiving lake water bodies. Lake depth had a negative effect on lake TP, and allowing the effect of lake depth to vary by region improved model fit and exhibited regional spatial patterns (Fig. 12b). Some of the regional variation in depth-effects on TP was related to a cross-scale interaction with regional baseflow. We found support for a positive cross-scale interaction between regional baseflow and lake depth effects on lake TP – in regions with high baseflow, lake depth had less of an effect on TP compared to regions with low regional baseflow (Fig. 13).

The top ranked TP models by lake connectivity type were similar to the ‘all lake type’ model with some differences (Appendix Tables C2 – C4). For Isolated lakes, including regional baseflow and cross-scale interactions only moderately increased the variance explained (Appendix Table C2). The top ranked model for Drainage-UPLK lakes did not include cross-scale interactions and regional baseflow was not related to TP based on 95% confidence intervals

(Appendix Table C4). Drainage lakes had the same predictor variables and cross-scale interaction as the all lake type model (Appendix Table C3).

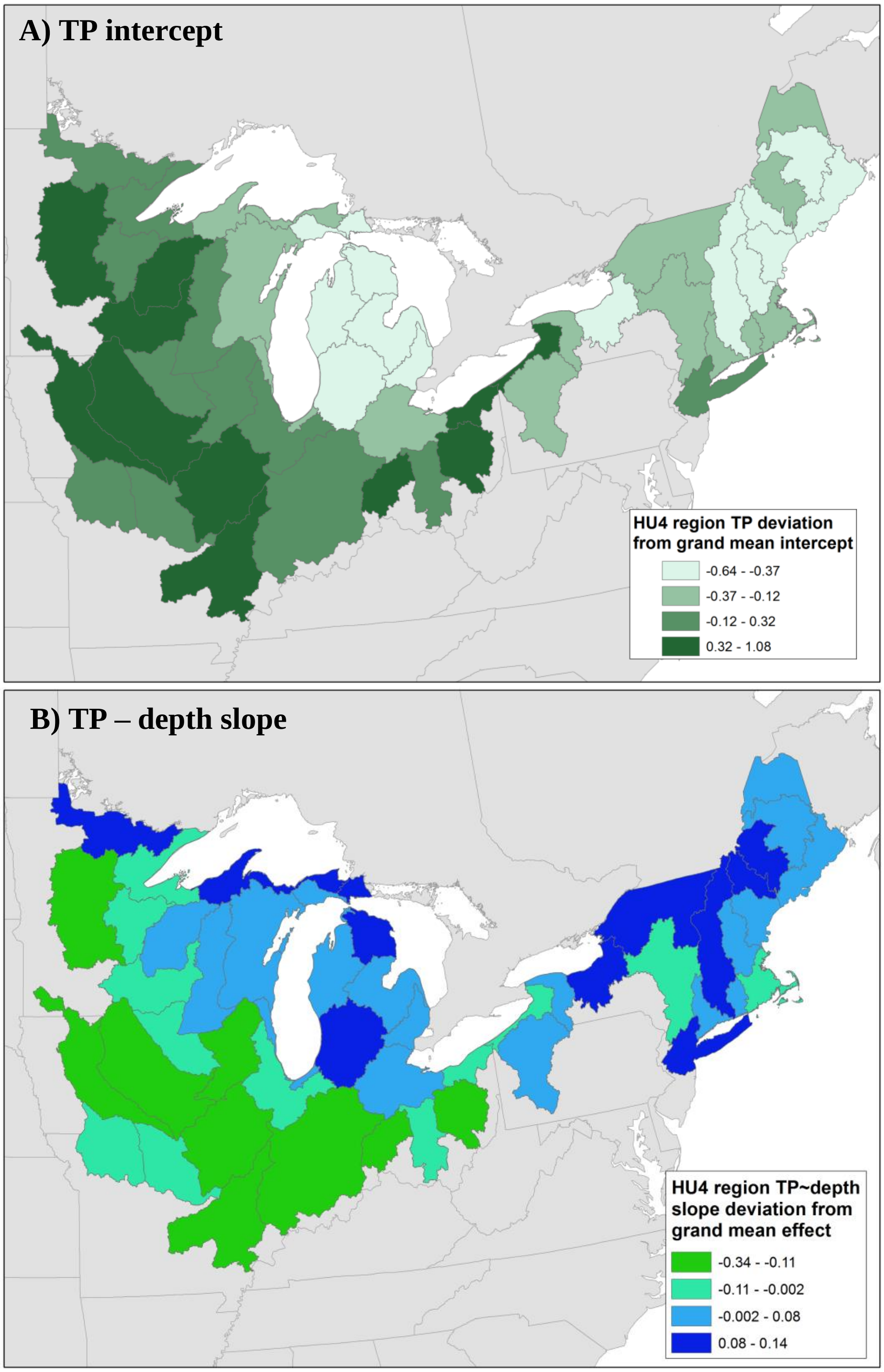

Figure 12.Deviation from the grand mean for A) TP intercepts, and B) local depth – TP slopes by HU4 region. Values near zero are close to the grand mean coefficient estimate.

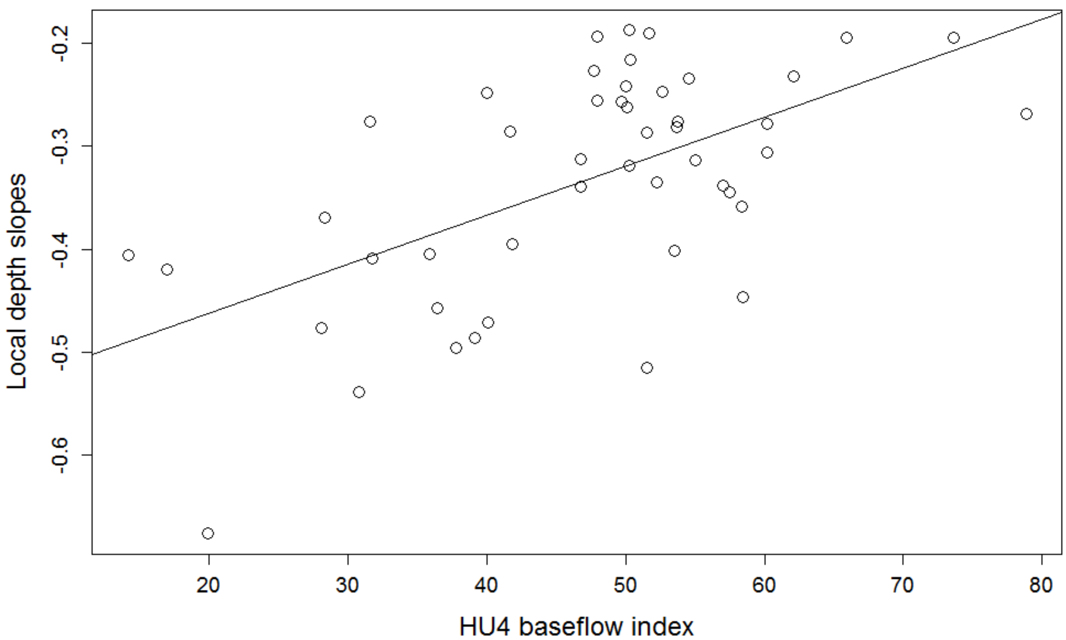

Figure 13. Cross-scale interaction between the effects of lake depth –on TP and regional baseflow. Lake depth slopes (open circles) represent lake depth relationships with lake TP by HU4 region. Lake depth was standardized prior to fitting the model. Baseflow, an indicator of groundwater, was measured in the HU4 region. Solid line is estimated hierarchical regression line.

Total nitrogen model

The top ranked TN model explained 65% of the total variation and only included local-scale predictor variables (Table 10). Average TN values varied across regions where TN values were higher than average in the Midwest and lower than average in the Northeast (Fig. 14a). Similar to the TP model, the TN model included internal lake processing – lake depth; nutrient sources – agriculture and urban land use; and freshwater transport – wetland cover that had similar effects. But, in contrast to the TP model, the top ranked TN model did not include any regional-scale variables or cross-scale interactions. At the local-scale, there was a positive interaction between wetlands in the watershed and lake depth to affect lake TN – in watersheds with high wetland cover, depth effects on TN were less negative. Model fit improved when wetland effects on lake TN were allowed to vary among regions. In the Northeast, wetlands tended to have a higher than average positive effect on TN (Fig. 14b). In general, regions where wetlands had less of a positive influence on TN tended to be in regions with higher than average TN. In our candidate models we did not identify any regional variables that were associated with TN and regional wetland effects. We did not include nitrogen deposition in the candidate models, which likely influences regional differences in lake TN. However, a large proportion of regional variation in lake TN was accounted for by local-scale variables suggesting that dominant nitrogen processes may be more related to local-scale variables that happen to be structured at regional scales.

The top ranked TN model differed by lake connectivity type (Appendix Tables C5 – C7). For Isolated lakes, drainage ratio was not important and Isolated wetlands (rather than total wetland) in the watershed were positively related to TN (Appendix Table C5). For Drainage lakes, the effects of local agriculture in the watershed on TN varied among regions; and agriculture had a positive interaction with lake depth such that in areas with high agriculture, lake depth had less negative effects on TN (Appendix Table C6). Drainage-UPLK lakes had similar predictor variables as the all lake type model (Appendix Table C7).

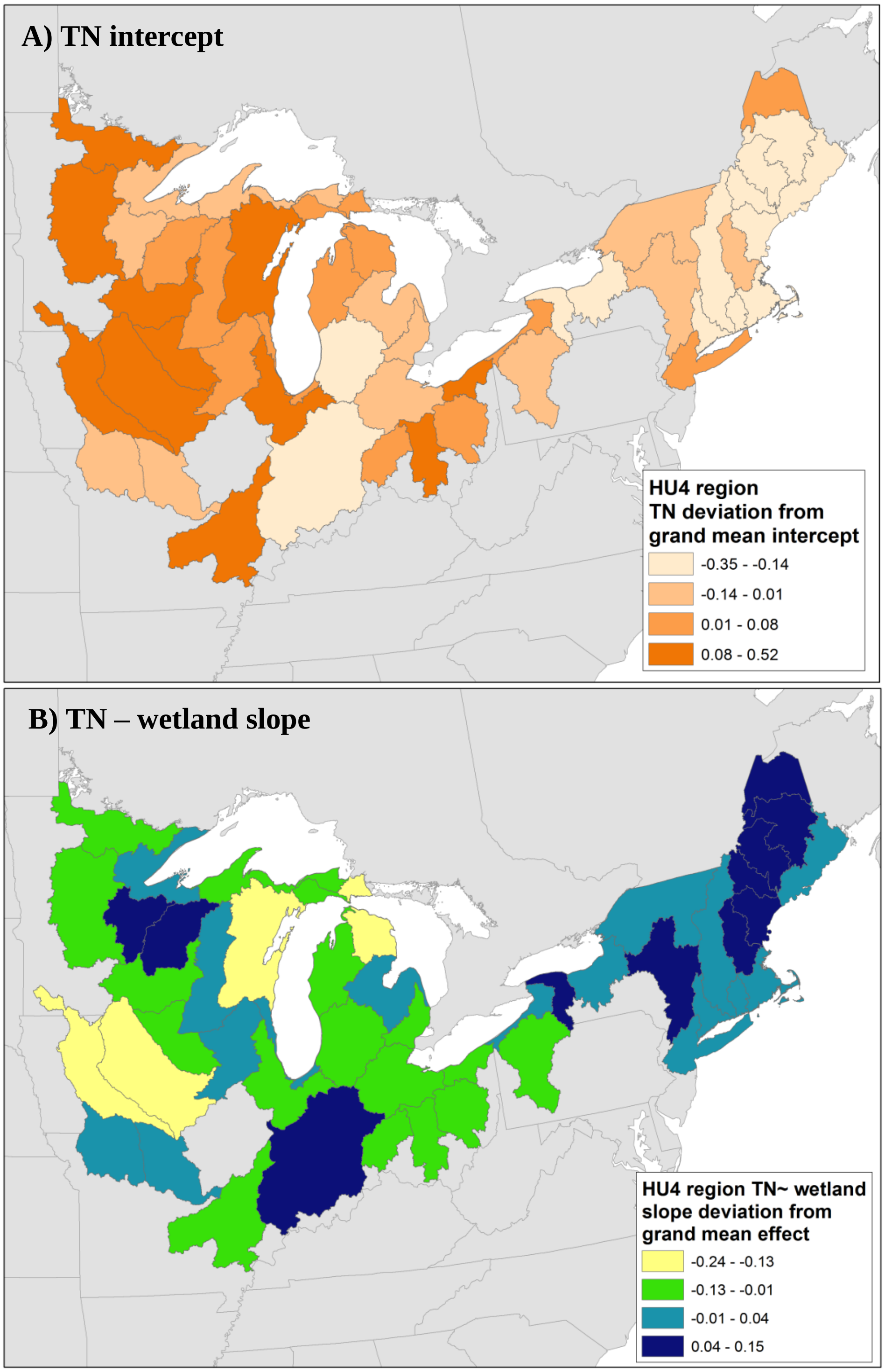

Figure 14. Deviation from the grand mean for A) TN intercepts, and B) local wetlands – TN slopes by HU4 region. Values near zero are close to the grand mean coefficient estimate.

Water color model

The top ranked model explained 38% of water color variation, the smallest amount of variation explained among the lake response variables (Table 11). Regions with higher than average water color were located in northern areas of the study extent (Fig. 15a). Local-scale predictor variables in the top ranked water color model represented similar mechanistic processes as in the TP and TN models. Lake depth, representing internal processing, was negatively related to water color. Wetlands in the watershed were positively related to water color and may capture both sources and transport processes of carbon. Wetland effects on water color varied by region. Some regions in the Northeast and upper Midwest had higher than average wetland effects on water color (Fig. 15b). Models that included regional forest and a cross-scale interaction between local wetland effects and regional forest were within four AIC units of the model with the lowest AIC value and can be interpreted as having a similar fit. The positive cross-scale interaction between local wetland effects and regional forest indicates that in regions with high forest cover, wetlands have a stronger positive effect on water color than in regions with lower forest cover.

The top ranked models among lake connectivity types had similar predictor variables as the all lake type model (Appendix Tables C8 – C10). Isolated lakes had the lowest variation explained by the predictor variables and wetland had less of a positive effect on water color in Isolated lakes compared to connected lake systems (i.e., Drainage and Drainage-UPLK lakes).

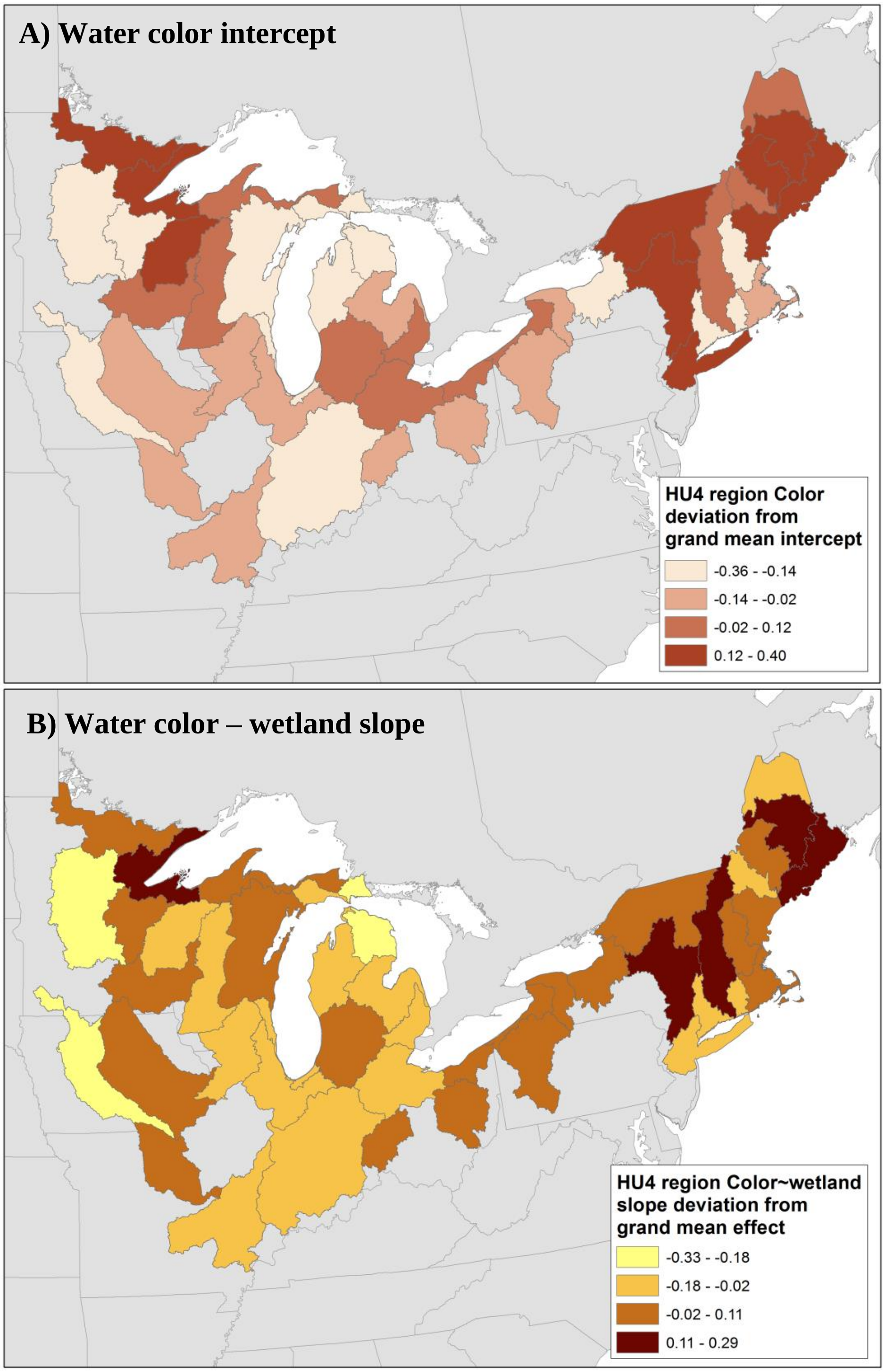

Figure 15.Deviation from the grand mean for A) water color intercepts, and B) local wetlands – color slopes by HU4 region. Values near zero are close to the grand mean coefficient estimate.

Discussion

We found that lake TP, TN, and water color varied regionally and were related to lake and landscape variables associated with internal processing, landscape nutrient and carbon sources, and freshwater transport. However, we also found that freshwater systems and their connectivity were related to lake water chemistry and were associated with both within- and cross-scale interactions for all three response variables. In particular, local wetlands in the watershed interacted with local lake depth and regional landscape features to affect lake TN and water color, respectively; whereas regional baseflow interacted with lake depth to affect lake TP. In addition, lake connectivity to streams and upstream lakes affected what variables were related to lake TP, TN, and water color, which highlights the importance of freshwater connectivity in influencing the types and scales of landscape predictors related to lake water chemistry.

One important finding from our study was the differences in regional variation between lake nutrients and carbon measures. Lake nutrients had a large proportion of variation at the regional scale compared to water color, suggesting that lake nutrients may be influenced by regional scale variables (or local scale variables structured at regional levels) compared to water color which may be driven more by local variables and processes. In contrast, in other regional settings than those we studied here, lake dissolved organic carbon can exhibit greater variation between regions than within regions (Seekell *et al.*, 2014); however, the spatial extent of the study spanned areas with significant latitudinal gradients that were not captured in our study extent. If we extended our temperate study extent into North American boreal regions, which are known to have highly colored lake systems (Prepas *et al.*, 2001), we would likely detect a greater regional signal in lake water color.

Synthesis of the relative importance of freshwater transport, sources, and internal processing of lake nutrients and carbon

Freshwater transport: We found that wetlands in the watershed were positively related to lake water chemistry but had a greater effect on water color compared to nutrients. Wetland effects on nutrients in surface waters have been shown to be highly variable from study to study (Detenbeck *et al.*, 1993; Johnson *et al.*, 1997; Devito *et al.*, 2000), but wetland effects on carbon measures have been fairly consistent across different regional settings (Xenopoulos *et al.*, 2003). For TP and TN, we found differences in the freshwater features that were related to lake nutrients, indicating that TP and TN may be affected by different freshwater landscape mechanisms. We found lake TP to be positively related to stream density in the watershed and negatively related to regional baseflow. Whereas we found lake TN to be positively related to only wetlands in the watershed.

Sources and internal processing*:* The local-scale predictors that were related to lake nutrients and carbon in the different models followed expected relationships with the underlying mechanistic processes they were associated with. Maximum lake depth was negatively associated with lake nutrients and water color such that deeper lakes had lower TP, TN, and water color. These relationships may represent internal processing and burial of nutrients and carbon in lakes (Kalff, 2002). Drainage ratio (CA:LK) was positively associated with nutrients and carbon, and lakes with high drainage ratios have been associated with short water retention time (Kalff, 2002; Webster *et al.*, 2008). Agriculture and urban land use in the watershed was positively associated with lake TP and TN, representing landscape sources of nutrients (KopciCek *et al.*, 1995; Carpenter *et al.*, 1998). And, wetland cover in the watershed was positively related to water color indicating a landscape source of carbon.

Local-scale interactions*:* Within local-scale interactions were only important in the TN model where wetlands in the watershed interacted with lake depth to affect lake TN. This positive interaction indicated that in watersheds with high wetland cover, lake depth effects were less negative on lake TN. However the effect size of this interaction was small and did not explain additional TN variation.

Regional predictors*:* The top ranked models among lake TP, TN, and water color did not always include a region-scaled predictor variable, despite the fact that there was regional variation in all three response variables as indicated by the ICC values. Lake phosphorus and water color were related to regional baseflow and forest cover, respectively. However, lake TN was not related to any regional predictor variables that we tested. It is likely that other regional-scaled variables that we did not include in our candidate models may be related to lake nutrients and water color. Nitrogen deposition would likely influence lake TN (Carpenter *et al.*, 1998). And climatic variables such as precipitation may affect the transport of nutrients and carbon to lake systems (Howarth *et al.*, 2006; Keller *et al.*, 2008). These regional variables warrant examination in future studies.

Cross-scale interactions*:* We found evidence of cross-scale interactions affecting lake TP and water color. For lake TP, there was a cross scale interaction where regional baseflow positively influenced lake depth effects on phosphorus. In regions with high baseflow, depth effects on phosphorus were less negative. Regional baseflow may directly influence lake depth effects on phosphorus by contributing groundwater to lakes that presumably would be less phosphorus enriched than surface water inflows due to the high adsorption of P to soil particles (Hayashi & Rosenberry, 2002). Alternatively, regional baseflow may be correlated with other regional-scale variables that affect lake depth effects (Appendix Table C11). In our study extent, regional baseflow was negatively correlated with regional stream density (r = -0.73) which may indicate that regions with high baseflow have fewer surface water flowpaths to deliver phosphorus to lakes and subsequently be processed within lakes. For lake water color, forest cover in the region positively influenced wetland effects on water color. Regional forest cover was negatively correlated with regional agriculture (r = -0.85) and in these less disturbed regions wetlands may receive carbon inputs from subsurface flows from forest vegetation. Lake TN was not related to a cross-scale interaction.

<u>The effect of lake type</u>*:* We found that Isolated lake models explained less variation in lake response variables compared to Drainage and Drainage-UPLK lake models. This pattern is not surprising and aligns with findings in the literature where landscape predictors for systems with small watersheds had less predictive power compared to systems with larger watersheds (Strayer *et al.*, 2003). Isolated lakes tend to have smaller watersheds in our study dataset, and the decrease in surface stream connections may make the spatial configuration of landscape features in relation to lakes highly important (Strayer *et al.*, 2003). This presents a challenge to develop accurate landscape models for Isolated lake systems when there is more variation and less explanatory power in the predictor variable relationships.

Evidence for controls of lake chemistry at macroscales

<u>TP:</u> Lake TP was influenced by similar local and regional variables in the separate lake connectivity type models. Across all lake types, lake depth effects on total phosphorus exhibited regional differences (i.e., random slope), and depth effects were more negative in regions where lake phosphorus concentrations were higher than average. These findings support results from a meta-analysis of lakes from various regions around the world that showed that depth effects on

lake TP were not fixed (Taranu & Gregory-Eaves, 2008). In addition, stream density in the watershed was positively related to lake TP across lake types, except for Isolated lakes. Streams transport both dissolved phosphorus and phosphorus bound to sediments across the landscape and can be a significant delivery pathways of phosphorus to receiving water bodies (Fraterrigo & Downing, 2008; Reed-Andersen *et al.*, 2014). We found that agricultural effects on phosphorus were greatest in Drainage-UPLK lakes, which aligns with findings from Nielsen *et al.* (2012) that lakes with streams in the watershed had stronger agriculture-TP effects compared to lakes without streams.

TN: The top ranked TN models were somewhat different among lake connectivity types. The transport mechanisms of nitrogen to lakes may be sensitive to freshwater connections as nitrogen is highly mobile in the landscape moving from sources to sinks through surface and subsurface flows (Howarth *et al.*, 2006). Lake-stream connectivity relationships likely influence these transport pathways (Fraterrigo & Downing, 2008). For Isolated and Drainage-UPLK lakes, wetland effects on lake TN exhibited regional differences (random slopes). Wetlands are recognized to play significant roles in nitrogen biogeochemical processing; and wetland morphology, hydrology, and biotic composition influence these processes (Jansson *et al.*, 1994; Mitsch & Gosselink, 2007). It is expected that wetland composition and hydrologic attributes exhibits regional variation which may promote variation in wetland effects on lake TN. For Drainage lakes, agriculture in the watershed exhibited different effects on lake TN by regions. These different effects may be picking up on regional differences in the type of agricultural activity occurring within the region which has been shown to export different N:P ratios (Arbuckle & Downing, 2001).

Color: Isolated lakes had similar predictor variables as the top ranked all lake type models and included a cross-scale interaction between regional forest and local wetlands effects on water color. In contrast, Drainage-UPLK lakes had the same predictor variables but there was no cross-scale interaction evident. These lakes may be somewhat buffered from regional landscape effects because upstream lakes may receive, process, and retain nutrients and carbon that would otherwise have been transported downstream. The Drainage lakes models did not converge with wetlands treated as a random slope and thus it is difficult to compare the model output with the other lake type models.

APPENDIX

Chapter 3: Appendix Tables and Figures

Table C1. Unconditional mixed-effects model intraclass correlation coefficient (ICC) by lake connectivity type. The ICC represents the percent of the total variation in lake response variable that occurs at the HU4 (region) scale. (n = number of lakes; N = number of HU4 regions)

	ICC %		
Lake type	TP	TN	Color
Isolated	50.00 (n = 1240; N = 40)	50.74 (n = 363; N =39)	10.11 (n = 430; N = 37)
Drainage	51.94 (n = 1838; N = 47)	54.66 (n=628; N =43)	9.10 (n = 546; N = 42)
Drainage-UPLK	53.97 (n = 1633; N = 44)	62.46 (n= 576; N = 40)	13.67 (n = 648; N = 38)

Table C2. Mixed-effects models predicting total phosphorus for *Isolated* lakes (TP; n = 1240 lakes, N = 40 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TP (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Models: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	2.86 (2.76, 3.11)	2.86 (2.72, 3.01)	2.86 (2.72, 2.99)	2.82 (2.70, 2.95)	2.81 (2.68, 2.94)
Internal process $_{Local}$	Depth	-0.30 (-0.34, -0.27)	-0.32 (-0.34, -0.27)	-0.35 (-0.42, -0.27)	-0.34 (-0.41, -0.26)	-0.33 (-0.40, -0.27)
	CA:LK	0.09 (0.03, 0.13)	0.09 (0.04, 0.14)	0.09 (0.03, 0.13)	0.09 (0.04, 0.13)	0.09 (0.04, 0.14)
Source $_{Local}$	Agriculture	0.29 (0.23, 0.33)	0.28 (0.24, 0.34)	0.27 (0.22, 0.31)	0.26 (0.21, 0.31)	0.26 (0.21, 0.31)
	Urban	0.19 (0.15, 0.22)	0.19 (0.15, 0.22)	0.18 (0.14, 0.22)	0.18 (0.14, 0.22)	0.18 (0.14, 0.21)
FW Connectivity $_{Local}$	Stream	0.09 (-0.04, 0.22)	–	–	–	–
	Wetland	0.09 (0.06, 0.12)	0.09 (0.06, 0.13)	0.09 (0.06, 0.12)	0.09 (0.06, 0.12)	0.09 (0.06, 0.12)
Source $_{Local}$ × Internal process $_{Local}$	Agriculture × Depth	–	-0.07 (-0.03, 0.06)	–	–	–
Random effects						
Internal process $_{random}$	Depth $_{random}$	–	–	Depth $_{random}$	Depth $_{random}$	Depth $_{random}$
Region-scale covariate						
FW Conn. $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.14 (-0.25, -0.03)	-0.21 (-0.34, -0.09)
Cross-scale interaction						
Internal process $_{Local}$ × FW Connectivity $_{Region}$	Depth $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	0.10 (0.02, 0.18)
Variance components	σ^2	0.34	0.34	0.32	0.32	0.32
	τ_{00}	0.16	0.19	0.15	0.12	0.12
	τ_{11}			0.03	0.03	0.02
Variation explained	Within region	28%	28%	32%	32%	32%
	Among region				74%	74%
	Total	47%	44%	50%	53%	53%
Model selection	AIC	2297	2295	2274	2274	2274
	Δ_i	24	22	0	0	0

Table C3. Mixed-effects models predicting total phosphorus for *Drainage* lakes (TP; n = 1838 lakes, N = 47 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TP (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

		Model				
Category	Variable	1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	3.04 (2.91, 3.18)	3.04 (2.91, 3.18)	3.04 (2.90, 3.18)	3.01 (2.87, 3.14)	2.93 (2.80, 3.05)
Internal process $_{Local}$	Depth	-0.32 (-0.35, -0.29)	-0.32 (-0.35, -0.29)	-0.37 (-0.43, -0.32)	-0.37 (-0.43, -0.31)	-0.33 (-0.37, -0.28)
	CA:LK	0.10 (0.06, 0.13)	0.10 (0.06, 0.13)	0.10 (0.06, 0.13)	0.10 (0.06, 0.13)	0.09 (0.06, 0.13)
Source $_{Local}$	Agriculture	0.37 (0.33, 0.41)	0.37 (0.33, 0.41)	0.36 (0.32, 0.40)	0.36 (0.32, 0.40)	0.35 (0.32, 0.39)
	Urban	0.19 (0.16, 0.23)	0.19 (0.16, 0.23)	0.19 (0.15, 0.22)	0.19 (0.15, 0.22)	0.19 (0.15, 0.22)
FW Connectivity $_{Local}$	Stream	0.06 (0.03, 0.09)	0.06 (0.03, 0.09)	0.06 (0.03, 0.09)	0.06 (0.02, 0.09)	0.06 (0.03, 0.09)
	Wetland	0.09 (0.05, 0.13)	0.09 (0.06, 0.14)	0.10 (0.06, 0.14)	0.10 (0.06, 0.14)	0.10 (0.06, 0.14)
Source $_{Local}$ × Internal process $_{Local}$	Agriculture × Depth	–	-0.05 (-0.08, -0.02)	0.01 (-0.02, 0.04)	–	–
Random effects						
Internal process $_{random}$	Depth $_{random}$	–	–	Depth $_{random}$	Depth $_{random}$	Depth $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.07 (-0.015, 0.02)	-0.23 (-0.33, -0.13)
Cross-scale interaction						
Internal process $_{Local}$ × FW Connectivity $_{Region}$	Depth $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	0.11 (0.07, 0.16)
Variance components	σ^2	0.34	0.35	0.33	0.33	0.33
	τ_{00}	0.20	0.20	0.20	0.17	0.14
	τ_{11}			0.02	0.02	0.01
Variation explained	Within region	38%	38%	41%	41%	41%
	Among region				72%	77%
	Total	53%	53%	54%	57%	59%
Model selection	AIC	3467	3464	3397	3402	3385
	Δ_i	82	79	12	17	0

Table C4. Mixed-effects models predicting total phosphorus for *Drainage-UPLK* lakes (TP; n = 1633 lakes, N = 44 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TP (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Model: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariates						
	Intercept	3.05 (2.90, 3.19)	3.05 (2.90, 3.19)	3.04 (2.90, 3.19)	3.01 (2.85, 3.15)	3.00 (2.85, 3.15)
Internal process $_{Local}$	Depth	-0.31 (-0.34, -0.28)	-0.31 (-0.34, -0.27)	-0.30 (-0.35, -0.26)	-0.30 (-0.35, -0.26)	-0.30 (-0.35, -0.26)
	CA:LK	0.10 (0.04, 0.13)	0.10 (0.07, 0.13)	0.11 (0.07, 0.14)	0.11 (0.07, 0.14)	0.11 (0.07, 0.14)
Source $_{Local}$	Agriculture	0.46 (0.41, 0.50)	0.47 (0.42, 0.51)	0.46 (0.41, 0.50)	0.45 (0.40, 0.50)	0.45 (0.40, 0.50)
	Urban	0.20 (0.16, 0.24)	0.20 (0.17, 0.24)	0.21 (0.17, 0.25)	0.21 (0.17, 0.24)	0.21 (0.17, 0.25)
FW Connectivity $_{Local}$	Stream	0.10 (0.06, 0.14)	0.10 (0.06, 0.14)	0.10 (0.05, 0.14)	0.10 (0.05, 0.14)	0.09 (0.05, 0.14)
	Wetland	0.10 (0.06, 0.13)	0.10 (0.07, 0.14)	0.10 (0.06, 0.13)	0.10 (0.06, 0.13)	0.10 (0.06, 0.13)
Source $_{Local}$ × Internal process $_{Local}$	Agriculture × Depth	–	-0.05 (-0.08, -0.02)	-0.04 (-0.07, -0.01)	-0.04 (-0.07, -0.01)	-0.04 (-0.07, -0.01)
Random effects						
Internal process $_{random}$	Depth $_{random}$	–	–	Depth $_{random}$	Depth $_{random}$	Depth $_{random}$
Region-scale covariates						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.11 (-0.23, 0.02)	-0.10 (-0.23, 0.03)
Cross-scale interaction						
Internal process $_{Local}$ × FW Connectivity $_{Region}$	Depth $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	-0.01 (-0.06, 0.04)
Variance components	σ^2	0.33	0.33	0.32	0.32	0.32
	τ_{00}	0.20	0.20	0.21	0.20	0.20
	τ_{11}			0.01	0.01	0.01
Variation explained	Within region	44%	44%	46%	46%	46%
	Among region				71%	71%
	Total	59%	59%	59%	59%	59%
Model selection	AIC	3003	3001	2990	2993	3001
	Δ_i	13	11	0	3	11

Table C5. Mixed-effects models predicting total nitrogen in *Isolated* lakes (TN; n = 363 lakes, N = 39 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TN (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i). Iso wet. = Isolated wetland

Category	Variable	Model: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	6.13 (5.99, 6.31)	6.12 (6.02, 6.22)	6.12 (6.02, 6.22)	6.13 (6.02, 6.22)	6.12 (6.01, 6.22)
Internal process $_{Local}$	Depth	-0.22 (-0.26, -0.18)	-0.24 (-0.28, -0.20)	-0.23 (-0.27, -0.19)	-0.23 (-0.27, -0.19)	-0.23 (-0.27, -0.19)
	CA:LK	-0.05 (-0.11, 0.01)	–	–	–	–
Source $_{Local}$	Agriculture	0.36 (0.30, 0.42)	0.30 (0.24, 0.37)	0.28 (0.21, 0.35)	0.29 (0.21, 0.35)	0.27 (0.20, 0.34)
	Urban	0.17 (0.12, 0.20)	0.17 (0.13, 0.21)	0.16 (0.12, 0.20)	0.16 (0.12, 0.20)	0.16 (0.12, 0.20)
FW Connectivity $_{Local}$	Stream	0.03 (-0.09, 0.20)	–	–	–	–
	Isolated Wetland	0.07 (0.04, 0.11)	0.08 (0.04, 0.11)	0.10 (0.04, 0.15)	0.09 (0.04, 0.15)	0.09 (0.04, 0.15)
Source $_{Local}$ × Internal process $_{Local}$	Agriculture × Depth	–	-0.10 (-0.08, -0.05)	-0.10 (-0.14, -0.05)	-0.10 (-0.14, -0.05)	-0.10 (-0.14, -0.05)
Random effects						
FW Connectivity $_{random}$	Iso wet. $_{random}$	–	–	Iso wet. $_{random}$	Iso wet. $_{random}$	Iso wet. $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.02 (-0.12, 0.08)	-0.02 (-0.12, 0.08)
Cross-scale interaction						
FW Connectivity $_{Local}$ × FW Connectivity $_{Region}$	Wetland $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	0.03 (-0.04, 0.10)
Variance components	σ^2	0.14	0.13	0.13	0.13	0.13
	τ_{00}	0.07	0.06	0.06	0.07	0.07
	τ_{11}			0.006	0.006	0.005
Variation explained	Within region	42%	46%	46%	46%	46%
	Among region	–	–	–	72%	72%
	Total	57%	61%	61%	59%	59%
Model selection	AIC	402	392	391	397	403
	Δ_i	11	1	0	6	12

Table C6. Mixed-effects models predicting total nitrogen in *Drainage* lakes (TN; n = 628 lakes, N = 43 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below.Variance components were estimated for within-region lake TN (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i). Agr. = Agriculture

Category	Variable	Model				
		1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	6.15 (6.06, 6.26)	6.15 (6.05, 6.26)	6.14 (6.06, 6.22)	6.12 (6.04, 6.21)	6.14 (6.05, 6.23)
Internal process $_{Local}$	Depth	-0.18 (-0.22, -0.14)	-0.18 (-0.22, -0.14)	-0.18 (-0.21, -0.14)	-0.18 (-0.22, -0.14)	-0.18 (-0.22, -0.14)
	CA:LK	0.12 (0.09, 0.18)	0.10 (0.06, 0.15)	0.10 (0.05, 0.14)	0.10 (0.05, 0.14)	0.10 (0.05, 0.14)
Source $_{Local}$	Agriculture	0.40 (0.35, 0.45)	0.41 (0.36, 0.46)	0.39 (0.32, 0.47)	0.37 (0.29, 0.45)	0.41 (0.33, 0.49)
	Urban	0.17 (0.12, 0.20)	0.16 (0.12, 0.20)	0.17 (0.13, 0.21)	0.17 (0.13, 0.21)	0.17 (0.13, 0.20)
FW Connectivity $_{Local}$	Stream	-0.03 (-0.07, 0.01)	–	–	–	–
	Wetland	0.16 (0.10, 0.20)	0.15 (0.10, 0.20)	0.15 (0.11, 0.20)	0.17 (0.12, 0.21)	0.17 (0.12, 0.21)
Source $_{Local}$ × Internal process $_{Local}$	Agriculture × Depth	–	0.07 (0.03, 0.11)	0.09 (0.06, 0.13)	0.09 (0.06, 0.13)	0.09 (0.05, 0.13)
Random effects						
Source $_{random}$	Agr. $_{random}$	–	–	Agr. $_{random}$	Agr. $_{random}$	Agr. $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.09 (-0.16, -0.01)	-0.07 (-0.15, 0.01)
Cross-scale interaction						
FW Connectivity $_{Local}$ × FW Connectivity $_{Region}$	Wetland $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	0.05 (0.01, 0.10)
Variance components	σ^2	0.16	0.16	0.15	0.15	0.15
	τ_{00}	0.08	0.08	0.05	0.05	0.06
	τ_{11}	–	–	0.02	0.02	0.02
Variation explained	Within region	43%	43%	46%	46%	46%
	Among region	–	–	–	85%	82%
	Total	61%	61%	68%	68%	66%
Model selection	AIC	751	745	717	719	723
	Δ_i	34	28	0	2	6

Table C7. Mixed-effects models predicting total nitrogen in *Drainage-UPLK* lakes (TN; n = 576 lakes, N = 40 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake TN (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Model: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	6.20 (6.10, 6.31)	6.21 (6.10, 6.31)	6.20 (6.10, 6.29)	6.19 (6.10, 6.29)	6.20 (6.10, 6.29)
Internal process $_{Local}$	Depth	-0.16 (-0.20, -0.13)	-0.18 (-0.21, -0.14)	-0.18 (-0.21, -0.14)	-0.18 (-0.21, -0.14)	-0.18 (-0.21, -0.14)
	CA:LK	0.06 (0.04, 0.10)	0.07 (0.04, 0.10)	0.06 (0.02, 0.09)	0.06 (0.02, 0.09)	0.06 (0.03, 0.09)
Source $_{Local}$	Agriculture	0.41 (0.35, 0.46)	0.40 (0.35, 0.46)	0.39 (0.34, 0.44)	0.38 (0.33, 0.44)	0.39 (0.33, 0.44)
	Urban	0.20 (0.16, 0.24)	0.20 (0.16, 0.24)	0.20 (0.16, 0.25)	0.20 (0.16, 0.24)	0.20 (0.16, 0.24)
FW Connectivity $_{Local}$	Stream	-0.02 (-0.06, 0.02)	–	–	–	–
	Wetland	0.12 (0.07, 0.16)	0.12 (0.08, 0.16)	0.07 (-0.02, 0.17)	0.09 (0.01, 0.19)	0.10 (0.01, 0.19)
FW Connectivity $_{Local}$ × Internal process $_{Local}$	Wetland × Depth	–	0.04 (0.02, 0.07)	0.05 (0.02, 0.08)	0.05 (0.02, 0.08)	0.05 (0.02, 0.08)
Random effects						
FW Connectivity $_{random}$	Wetland $_{random}$	–	–	Wetland $_{random}$	Wetland $_{random}$	Wetland $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.08 (-0.17, 0.01)	-0.08 (-0.17, 0.01
Cross-scale interaction						
FW Connectivity $_{Local}$ × FW Connectivity $_{Region}$	Wetland $_{Local}$ × Baseflow $_{Region}$	–	–	–	–	-0.02 (-0.11, 0.07)
Variance components	σ^2	0.14	0.14	0.13	0.13	0.13
	τ_{00}	0.08	0.08	0.05	0.06	0.06
	τ_{11}	–	–	0.05	0.04	0.04
Variation explained	Within region	39%	39%	43%	43%	43%
	Among region	–	–	–	84%	84%
	Total	64%	64%	70%	69%	69%
Model selection	AIC	618	618	607	610	617
	Δ_i	11	11	0	3	10

Table C8. Mixed-effects models predicting water color for *Isolated* lakes (Color; n = 430 lakes, N = 37 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake Color (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Model: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	2.67 (2.46, 2.86)	2.68 (2.56, 2.80)	2.68 (2.55, 2.80)	2.68 (2.57, 2.80)	2.70 (2.59, 2.80)
Internal process $_{Local}$	Depth	-0.24 (-0.31, -0.17)	-0.24 (-0.31, -0.17)	-0.23 (-0.30, -0.17)	-0.23 (-0.30, -0.16)	-0.23 (-0.30, -0.16)
	CA:LK	0.23 (0.14, 0.32)	0.23 (0.14, 0.32)	0.23 (0.15, 0.32)	0.23 (0.14, 0.31)	0.23 (0.15, 0.32)
Source $_{Local}$	Forest	0.03 (-0.05, 0.10)	–	–	–	–
FW Connectivity $_{Local}$	Stream	-0.01 (-0.21, 0.18)	–	–	–	–
	Wetland	0.25 (0.20, 0.31)	0.27 (0.20, 0.33)	0.26 (0.18, 0.35)	0.26 (0.17, 0.34)	0.27 (0.20, 0.34)
FW Connectivity $_{Local}$ × Internal process $_{Local}$	Wetland × Depth	–	0.02 (-0.03, 0.07)	–	–	–
Random effect						
FW Connectivity $_{random}$	Wetland $_{random}$	–	–	Wetland $_{random}$	Wetland $_{random}$	Wetland $_{random}$
Region-scale covariate						
Source $_{Region}$	Forest $_{Region}$	–	–	–	0.10 (0.01, 0.19)	0.15 (0.06. 0.25)
Cross-scale interaction						
FW Connectivity $_{Local}$ × Source $_{Region}$	Wetland $_{Local}$ × Forest $_{Region}$	–	–	–	–	0.11 (0.04, 0.18)
Variance components	σ^2	0.39	0.40	0.38	0.38	0.38
	τ_{00}	0.05	0.05	0.05	0.05	0.04
	τ_{11}	–	–	0.02	0.02	0.02
Variation explained	Within region	32%	30%	33%	33%	33%
	Among region	–	–	–	17%	17%
	Total	30%	29%	32%	32%	33%
Model selection	AIC	880	887	873	875	875
	Δ_i	7	14	0	2	2

Table C9. Mixed-effects models predicting water color for *Drainage* lakes (Color; n = 546 lakes, N = 42 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake Color (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

		Model				
Category	Variable	1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	2.81 (2.72, 2.90)	2.82 (2.71, 2.91)	2.80 (2.71, 2.89)	2.80 (2.70, 2.88)	2.79 (2.70, 2.88)
Internal process $_{Local}$	Depth	-0.23 (-0.29, -0.17)	-0.22 (-0.28, -0.16)	-0.23 (-0.29, -0.17)	-0.23 (-0.29, -0.17)	-0.23 (-0.29, -0.17)
	CA:LK	0.33 (0.25, 0.40)	0.33 (0.25, 0.39)	0.34 (0.27, 0.41)	0.34 (0.27, 0.41)	0.34 (0.27, 0.41)
Source $_{Local}$	Forest	0.07 (-0.01, 0.13)	–	0.05 (-0.02, 0.13)	0.06 (-0.01, 0.14)	0.07 (-0.01, 0.15)
FW Connectivity $_{Local}$	Stream	0.02 (-0.04, 0.08)	–	–	–	–
	Wetland	0.47 (0.40, 0.54)	0.48 (0.39, 0.55)	0.47 (0.40, 0.54)	0.48 (0.40, 0.55)	0.48 (0.40, 0.55)
FW Connectivity $_{Local}$ × Internal process $_{Local}$	Wetland × Depth	–	0.02 (-0.04, 0.07)	–	–	–
Random effect						
Source $_{random}$	Forest $_{random}$	–	–	Forest $_{random}$	Forest $_{random}$	Forest $_{random}$
Region-scale covariate						
FW Connectivity $_{Region}$	Baseflow $_{Region}$	–	–	–	-0.04 (-0.10, 0.03)	-0.03 (-0.12, 0.06)
Cross-scale interaction						
FW Connectivity $_{Local}$ × Source $_{Region}$	Wetland $_{Local}$ × Forest $_{Region}$	–	–	–	–	0.01 (-0.06, 0.08)
Variance components	σ^2	0.37	0.39	0.37	0.37	0.38
	τ_{00}	0.04	0.05	0.03	0.03	0.04
	τ_{11}	–	–	0.01	0.01	0.02
Variation explained	Within region	44%	41%	44%	44%	42%
	Among region	–	–	–	57%	43%
	Total	44%	40%	45%	45%	42%
Model selection	AIC	1076	1083	1074	1080	1087
	Δ_i	2	9	0	6	13

Table C10. Mixed-effects models predicting water color for *Drainage-UPLK* lakes (Color; n = 648 lakes, N = 38 regions). Model coefficients and 95% confidence intervals (lower, upper) are provided below. Variance components were estimated for within-region lake Color (σ^2), among-region intercept (τ_{00}), and among-region covariate slopes (τ_{11}). Candidate models (1 – 5) were compared using Akaike Information Criteria (AIC) by taking the difference in AIC values with the lowest AIC model (Δ_i).

Category	Variable	Model: 1. Local	2. Local interactions	3. Random slopes	4. Region	5. CSI
Local-scale covariate						
	Intercept	2.88 (2.79, 2.99)	2.87 (2.78, 2.97)	2.89 (2.79, 2.99)	2.89 (2.79, 2.99)	2.89 (2.79, 2.99)
Internal process $_{Local}$	Depth	-0.20 (-0.25, -0.15)	-0.20 (-0.25, -0.15)	-0.20 (-0.25, -0.15)	-0.20 (-0.25, -0.15)	-0.20 (-0.25, -0.15)
	CA:LK	0.21 (0.17, 0.27)	0.21 (0.17, 0.26)	0.21 (0.16, 0.25)	0.21 (0.16, 0.25)	0.21 (0.16, 0.25)
Source $_{Local}$	Forest	0.01 (-0.06, 0.08)	–	–	–	–
FW Connectivity $_{Local}$	Stream	-0.03 (-0.09, 0.03)	–	–	–	–
	Wetland	0.30 (0.24, 0.36)	0.30 (0.24, 0.36)	0.32 (0.20, 0.45)	0.32 (0.21, 0.45)	0.33 (0.21, 0.45)
FW Connectivity $_{Local}$ × Internal process $_{Local}$	Wetland × Depth	–	-0.01 (-0.06, 0.05)	–	–	–
Random effect						
FW Connectivity $_{random}$	Wetland $_{random}$	–	–	Wetland $_{random}$	Wetland $_{random}$	Wetland $_{random}$
Region-scale covariate						
Source $_{Region}$	Forest $_{Region}$	–	–	–	0.03 (-0.05, 0.12)	0.04 (-0.05, 0.13)
Cross-scale interaction						
FW Connectivity $_{Local}$ × Source $_{Region}$	Wetland $_{Local}$ × Forest $_{Region}$	–	–	–	–	0.03 (-0.08, 0.15)
Variance components	σ^2	0.37	0.37	0.35	0.35	0.35
	τ_{00}	0.05	0.05	0.04	0.04	0.05
	τ_{11}	–	–	0.06	0.06	0.06
Variation explained	Within region	34%	34%	38%	38%	38%
	Among region	–	–	–	56%	44%
	Total	35%	35%	40%	40%	33%
Model selection	AIC	1265	1272	1249	1255	1260
	Δ_i	16	23	0	6	11

Table C11. Correlation coefficients of local lake and watershed and regional (HU4) predictor variables.

	Depth	CA:LK	Wet %	Agr %	Urb %	Stream	Wet $_{HU4}$	Agr $_{HU4}$	Urb $_{HU4}$	For $_{HU4}$	Strm $_{HU4}$	Base $_{HU4}$
Depth	1											
CA:LK	-0.22	1										
Wet %	-0.11	-0.03	1									
Agr %	0.02	0.25	-0.09	1								
Urb %	-0.10	-0.07	-0.11	-0.04	1							
Stream	-0.01	0.45	-0.11	0.15	-0.07	1						
Wet $_{HU4}$	0.06	-0.19	0.38	-0.13	-0.09	-0.39	1					
Agr $_{HU4}$	0.03	0.07	0.01	0.69	0.17	-0.02	-0.16	1				
Urb $_{HU4}$	-0.19	0.10	-0.09	0.03	0.47	0.08	-0.35	0.08	1			
For $_{HU4}$	-0.01	-0.02	-0.11	-0.59	-0.31	0.12	-0.09	-0.85	-0.32	1		
Strm $_{HU4}$	-0.08	0.16	-0.32	0.12	0.04	0.41	-0.76	0.20	0.22	0.10	1	
Base $_{HU4}$	-0.01	-0.09	0.28	-0.23	-0.03	-0.32	0.58	-0.33	-0.08	0.08	-0.73	1

LITERATURE CITED

LITERATURE CITED

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**, 970–975.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014) Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat]*.

Butman, D. & Raymond, P.A. (2011) Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geoscience*, **4**, 839–842.

Carpenter, S.R., Caraco, N.F., Correll, D.L., Howarth, R.W., Sharpley, A.N. & Smith, V.H. (1998) Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, **8**, 559–568.

Cheruvelil, K.S., Soranno, P.A., Webster, K.E. & Bremigan, M.T. (2013) Multi-scaled drivers of ecosystem state: quantifying the importance of the regional spatial scale. *Ecological Applications*, **23**, 1603–1618.

D'Arcy, P. & Carignan, R. (1997) Influence of catchment topography on water chemistry in southeastern Québec Shield lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 2215–2227.

Detenbeck, N.E., Johnston, C.A. & Niemi, G.J. (1993) Wetland effects on lake water quality in the Minneapolis/St. Paul metropolitan area. *Landscape Ecology*, **8**, 39–61.

Devito, K.J., Creed, I.F., Rothwell, R.L. & Prepas, E.E. (2000) Landscape controls on phosphorus loading to boreal lakes: implications for the potential impacts of forest harvesting. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 1977–1984.

Fergus, C.E., Soranno, P.A., Cheruvelil, K.S. & Bremigan, M.T. (2011) Multiscale landscape and wetland drivers of lake total phosphorus and water color. *Limnology and Oceanography*, **56**, 2127–2146.

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**, 1691–1703.

Fraterrigo, J.M. & Downing, J.A. (2008) The Influence of Land Use on Lake Nutrients Varies with Watershed Transport Capacity. *Ecosystems*, **11**, 1021–1034.

Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Hayashi, M. & Rosenberry, D.O. (2002) Effects of Ground Water Exchange on the Hydrology and Ecology of Surface Water. *Ground Water*, **40**, 309–316.

Haygarth, P.M., Condron, L.M., Heathwaite, A.L., Turner, B.L. & Harris, G.P. (2005) The phosphorus transfer continuum: Linking source to impact with an interdisciplinary and multi-scaled approach. *Science of The Total Environment*, **344**, 5–14.

Howarth, R.W., Swaney, D.P., Boyer, E.W., Marino, R., Jaworski, N. & Goodale, C. (2006) The influence of climate on average nitrogen export from large watersheds in the Northeastern United States. *Biogeochemistry*, **79**, 163–186.

Jansson, M., Andersson, R., Berggren, H. & Leonardson, L. (1994) Wetlands and Lakes as Nitrogen Traps. *Ambio*, **23**, 320–325.

J. Pinheiro, D. Bates, S. DebRoy & D. Sarkar (2016) *nlme: Linear and nonlinear mixed effects models*,.

Johnson, L., Richards, C., Host, G. & Arthur, J. (1997) Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology*, **37**, 193–208.

Kalff, J.A. (2002) *Limnology: Inland water ecosystems*, Prentice Hall.

Keller, W. (Bill), Paterson, A.M., Somers, K.M., Dillon, P.J., Heneberry, J. & Ford, A. (2008) Relationships between dissolved organic carbon concentrations, weather, and acidification in small Boreal Shield lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 786–795.

KopciCek, J., Prochdzkovci, L., Stuchlik, E. & Bla2ka, P. (1995) The nitrogen phosphorus relationship in mountain lakes: Influence of atmospheric input, watershed, and pH. *Limnology and Oceanography*, **40**, 930–937.

Laudon, H., Berggren, M., Ågren, A., Buffam, I., Bishop, K., Grabs, T., Jansson, M. & Köhler, S. (2011) Patterns and Dynamics of Dissolved Organic Carbon (DOC) in Boreal Streams: The Role of Processes, Connectivity, and Scaling. *Ecosystems*, **14**, 880–893.

Meybeck, M. (1995) *Global Distribution of Lakes*. *Physics and Chemistry of Lakes* (ed. by P.A. Lerman), P.D.D.M. Imboden), and P.J.R. Gat), pp. 1–35. Springer Berlin Heidelberg.

Mitsch, W.J. & Gosselink, J.G. (2007) *Wetlands*, John Wiley & Sons.

Nielsen, A., Trolle, D., Søndergaard, M., Lauridsen, T.L., Bjerring, R., Olesen, J.E. & Jeppesen, E. (2012) Watershed land use effects on lake water quality in Denmark. *Ecological Applications*, **22**, 1187–1200.

Peters, D.P., Groffman, P.M., Nadelhoffer, K.J., Grimm, N.B., Collins, S.L., Michener, W.K. & Huston, M.A. (2008) Living in an increasingly connected world: a framework for continental-scale environmental science. *Frontiers in Ecology and the Environment*, **6**, 229–237.

Peters, D.P.C., Bestelmeyer, B.T. & Turner, M.G. (2007) Cross–Scale Interactions and Changing Pattern–Process Relationships: Consequences for System Dynamics. *Ecosystems*, **10**, 790–796.

Prepas, E.E., Planas, D., Gibson, J.J., Vitt, D.H., Prowse, T.D., Dinsmore, W.P., Halsey, L.A., McEachern, P.M., Paquet, S., Scrimgeour, G.J., Tonn, W.M., Paszkowski, C.A. & Wolfstein, K. (2001) Landscape variables influencing nutrients and phytoplankton communities in Boreal Plain lakes of northern Alberta: a comparison of wetland- and upland-dominated catchments. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1286–1299.

Rantakari, M., Kortelainen, P., Vuorenmaa, J., Mannio, J. & Forsius, M. (2004) *Finnish Lake Survey: The Role of Catchment Attributes in Determining Nitrogen, Phosphorus, and Organic Carbon Concentrations*. *Biogeochemical Investigations of Terrestrial, Freshwater, and Wetland Ecosystems across the Globe* (ed. by R.K. Wieder), M. Novák), and M.A. Vile), pp. 683–699. Springer Netherlands.

Rasmussen, J.B., Godbout, L. & Schallenberg, M. (1989) The humic content of lake water and its relationship to watershed and lake morphometry. *Limnology and Oceanography*, **34**, 1336–1343.

Read, E.K., Patil, V.P., Oliver, S.K., Hetherington, A.L., Brentrup, J.A., Zwart, J.A., Winters, K.M., Corman, J.R., Nodine, E.R., Woolway, R.I., Dugan, H.A., Jaimes, A., Santoso, A.B., Hong, G.S., Winslow, L.A., Hanson, P.C. & Weathers, K.C. (2015) The importance of lake-specific characteristics for water quality across the continental United States. *Ecological Applications*, **25**, 943–955.

Reed-Andersen, T., Carpenter, S.R. & Lathrop, R.C. (2014) Phosphorus Flow in a Watershed-Lake Ecosystem. *Ecosystems*, **3**, 561–573.

Seekell, D.A., Lapierre, J.-F., Pace, M.L., Gudasz, C., Sobek, S. & Tranvik, L.J. (2014) Regional-scale variation of dissolved organic carbon concentrations in Swedish lakes. *Limnology and Oceanography*, **59**, 1612–1620.

Seitzinger, S., Harrison, J.A., Böhlke, J.K., Bouwman, A.F., Lowrance, R., Peterson, B., Tobias, C. & Drecht, G.V. (2006) Denitrification Across Landscapes and Waterscapes: A Synthesis. *Ecological Applications*, **16**, 2064–2090.

Smith, S.V., Renwick, W.H., Bartley, J.D. & Buddemeier, R.W. (2002) Distribution and significance of small, artificial water bodies across the United States landscape. *Science of The Total Environment*, **299**, 21–36.

Sobek, S., Tranvik, L.J., Prairie, Y.T., Kortelainen, P. & Cole, J.J. (2007) Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. *Limnology and Oceanography*, **52**, 1208–1219.

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

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

Soranno, P.A., Webster, K.E., Cheruvelil, K.S. & Bremigan, M.T. (2009) The lake landscape-context framework: linking aquatic connections, terrestrial features and human effects at multiple spatial scales. *Verhandlungen der Internationalen Vereinigung fur Theoretische und Angewandte Limnologie*, **30**, 695–700.

Strayer, D.L., Beighley, R.E., Thompson, L.C., Brooks, S., Nilsson, C., Pinay, G. & Naiman, R.J. (2003) Effects of Land Cover on Stream Ecosystems: Roles of Empirical Models and Scaling Issues. *Ecosystems*, **6**, 407–423.

Taranu, Z.E. & Gregory-Eaves, I. (2008) Quantifying Relationships Among Phosphorus, Agriculture, and Lake Depth at an Inter-Regional Scale. *Ecosystems*, **11**, 715–725.

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

Webster, K.E., Soranno, P.A., Cheruvelil, K.S., Bremigan, M.T., Downing, J.A., Vaux, P.D., Asplund, T.R., Bacon, L.C. & Connor, J. (2008) An empirical evaluation of the nutrient-color paradigm for lakes. *Limnology and Oceanography*, **53**, 1137–1148.

Wetzel, R.G. (2001) *Limnology: Lake and River Ecosystems*, Academic Press.

Williamson, C.E., Morris, D.P., Pace, M.L. & Olson, O.G. (1999) Dissolved organic carbon and nutrients as regulators of lake ecosystems: Resurrection of a more integrated paradigm. *Limnology and Oceanography*, **44**, 795–803.

Xenopoulos, M.A., Lodge, D.M., Frentress, J., Kreps, T.A., Bridgham, S.D., Grossman, E. & Jackson, C.J. (2003) Regional comparisons of watershed determinants of dissolved organic carbon in temperate lakes from the Upper Great Lakes region and selected regions globally. *Limnology and Oceanography*, **48**, 2321–2334.