Does Lake and Stream Connectivity Control Phosphorus Retention in Lakes?

Abstract

Lake water residence time and depth are known to be strong predictors of phosphorus (P) retention. However, there is substantial variation in P retention among lakes with the same depth and residence time. One potential explanatory factor for this remaining variation is that connectivity among lakes as well as connectivity between lakes and nearby streams influences either lake P trapping or the type of stream delivered P. Therefore, we examined the extent to which connectivity among lakes versus connectivity between lakes and streams contributes to differences in P retention among lakes. Specifically, we evaluated the effect of lake-based and stream-based connectivity metrics on P retention using a hierarchical parameterization of the Vollenweider equation. We compared Vollenweider's k in lakes with high and low stream versus lake connectivity. We found that variation in k is more strongly associated with lake connectivity metrics compared to stream connectivity metrics. This result suggests that lake-associated processes, which likely control P trapping, play a larger role in determining P retention than stream-associated processes which may affect variation in the type of delivered P.

Introduction

Importance of nutrient predictions

A comprehensive understanding of P cycling at broad scales is necessary to predict nutrient concentrations among many different lake types and to better manage the risks of eutrophication from excess nutrient loading (citation). Such predictions can ultimately be used to implement management actions aimed at avoiding phenomena such as algal blooms and fish kills which can negatively affect human health and degrade ecosystem services (Smith 2003).

Importance of P retention

Although both nitrogen (N) and phosphorus loading contribute to eutrophication risk, studies on lake eutrophication commonly focus solely on phosphorus (P). This focus on P is a result of strong evidence that P rather than N is the primary limiting nutrient in lakes (Schindler 1977). Although later studies found that N and N:P ratios strongly control primary production in some lakes, P concentration is strongly correlated with phytoplankton primary production in most lakes (Downing et al. 2001, Elser et al. 2007). In the present study, we focus on P retention rather than P concentration because it is a unitless (%)

measure that can be compared among different lakes irrespective of their baseline nutrient concentrations or total nutrient inputs (Brett and Benjamin 2007). P retention is typically bounded at 0 in the case where all P inputs exit the lake via outflow or sedimentation and bounded at 1 in the case where all nutrient inputs are retained in the water column (citation).

Knowledge Gap

One of the primary challenges of predicting P retention is that lake nutrient concentrations are not simply a function of total inputs to the lake, but rather a function of complex interactions between lake characteristics and biogeochemical cycles. For example, lake phosphorus concentration is a function of biological response to phosphorus loading yet this response is not only a function of load but also depends on lake characteristics such as depth and biogeochemical cycling with iron (Sondergaard 2003, Wagner et al. 2011). Such interactions may obscure direct relationships between nutrient concentrations and total nutrient inputs and ultimately prevent accurate prediction of lake nutrient concentrations.

One way to account for these complex interactions between lake characteristics and biogeochemical cycles is to undertake detailed processs-based study of each indvidual component that influences lake P processing (citations). Specific components may include things like species specific-phytoplankton settling rates, POM speciation, or detailed representation of elemental cycling at the sediment-water interface (citations). Clearly, such detailed process-based study may not be feasible for hundreds of lakes at the regional scale and may be difficult to use for general understanding of variation in P retention among many different lake types.

Expectations/Hypothesis

An alternative approach for predicting lake nutrient concentrations is static mass-balance modelling using techniques such as the Vollenweider equations which typically predict P concentrations as a function of P loading and water residence time (Vollenweider 1975). Although such an approach cannot represent the complexity of potential interactions among lake characteristics and biogeochemical cycles, it offers the ability to evaluate relationships between P retention and lake characteristics among make lakes at the regional scale. In the present study, we extend the Vollenweider approach to explore the possibility that connectivity may explain some of the variation in P retention among lakes with the same depth and residence time. We expect that connectivity among lakes as well as connectivity between lakes and nearby streams may indicate differences in either watershed P trapping or the type of stream delivered P. Specifically, we tested the hypothesis that lake associated processes play a larger role in determining P retention than stream associated processes.

Methods

Dataset description

We used P retention and residence time data from the National Eutrophication Survey (USEPA 1975, Stachelek et al. 2018) for lakes located in 17 northeastern-most US states. Values in the dataset represent mean summer measurements spanning a period of 1972 to 1975. Lakes were excluded from analysis if they had a surface area of greater than 1000 km2 or less than 0.1 km2 and excluded if they had a depth of greater than 70 m. Lakes were also excluded if they lacked any upstream surface water connections or were located on the coast (e.g. had an ocean or inland sea in the upstream watershed). Overall, this left a dataset of approximately 130 lakes.

Empirical models

We modelled lake P retention using the Vollenweider equations (Vollenweider 1975) which relate P retention to water residence time. Although there are several variants of these equations, we limited our analysis to a specific 2-parameter form that has been shown to have good performance in multiple cross-sectional studies (Brett and Benjamin 2007, Cheng et al. 2010):

$$R_i = 1 - \frac{1}{1 + k\tau_i^x} \tag{1}$$

where R_i is the P retention in lake i, k and x are adjustable (fitted) parameter coefficients, and τ is water residence time in lake i. The value of k has been variously interpreted as either a bulk loss rate or as an approximation of sedimentation rate (citation). We first used Eq. 1 to model the overall (non-hierarchical) relationship between P retention and water residence time for all lakes in the study dataset. Next, we explored hierarchical versions of Eq. 1 where is k modelled seperately k_j as a function of a sub-population indicator g_i denoting membership in one of two groups (described below):

$$R_i = 1 - \frac{1}{1 + k_j\tau_i^x} \tag{2}$$

$$k_j = g_i \tag{3}$$

We assigned lake membership in each sub-population group based on several lake characteristics and connectivity metrics where the splitting criteria was determined from the results of conditional inference trees (citation). The conditional inference tree (ctree) technique creates binary splits of the independent variables, which are recursively repeated, to find the split that maximizes association with

the depededent variable (citation). In the present study, our dependent variable was P retention while our independent variable was a single lake characteristics or connectivity metric (Table 1). The advantage of the ctree technique over the more widely used classification-regression tree (CART) technique is that tree growth stopping rules are pre-specified. As a result, this avoids some of the subjectivity associated with post-hoc tree pruning.

We fit both non-hierarchical and hierarchical models in a Bayesian framework using the non-linear extension to the `brms` package to interface with the Stan statistical program (Burkner 2017, Stan Development Team 2017). In both models, we set an informative prior on k and x of N(1.3, 0.1) and N(0.45, 0.1) respectively. These priors were based on the confidence intervals presented in Brett and Benjamin (2007) and qualitatively matched those used by Cheng et al. (2010). We used the default settings of `brms` and `rstan` to generate posterior estimates using four chains of 4,000 iterations each with no thinning and random intial parameter values between -2 and 2.

Connectivity metrics

We calculated several connectivity metrics in order to fit our hierarchical versions of the Vollenweider equations (Eq. 2) and to ultimately determine the importance of stream and lake connectivity as they relate to P retention. Our stream connectivity metrics included average link length, stream order ratio, and the closest distance to an upstream lake. We focused on average link length and stream order ratio because they are approximations of stream network complexity (citation, fractal stream book). For example, we consider low connectivity lakes to be those with highly complex stream networks characterized by a short average link length and a high stream order ratio. Our first step in calculating average link length was to dissolve stream network nodes that do not occur at a stream junction. Next, we divided the total stream length in the upstream watershed by the total number of stream reaches. We calculated (Strahler) stream order ratio by dividing the number of first-order streams in the upstream watershed of a given lake by the total number of higher order (> 1) streams (citation). Finally, we calculated closest distance to an upstream lake on a path-distance basis (as the fish swims) rather than on a straight-line basis (as the crow flies).

In addition to stream connectivity metrics, we calculated several lake connectivity metrics including total upstream lake area, number of upstream lakes, and presence/abscence of an upstream lake. For the purposes of calculating lake connectivity metrics, we defined a lake as any waterbody with an area greater than 4 ha (0.04 km2). In terms of lake connectivity, we consider a low connectivity lake to be one with low total upstream lake area and few upstream lakes.

Finally, we fit hierarchical models on the basis of partitions of several non-connectivity lake characteristics including maximum depth, baseflow, and wetland cover. This was intended as a check on the relative importance of connectivity

effects relative to lake characteristics known to influence P retention such as maximum depth.

We calculated both lake and stream connectivity metrics at multiple scales (Figure 1). First, we calculated connectivity metrics at the scale of individual lake watersheds (iws scale). We defined an individual lake watershed as the area draining to a particular lake exclusive of any upstream areas that drain into a lake greater than or equal to 10 ha (0.1 km2). We also calculated connectivity metrics at the scale of entire upstream lake networks (nws scale). Network watersheds are defined as the area draining into any part the upstream network irrespective of the presence or abscence of upstream lakes (Figure 1).
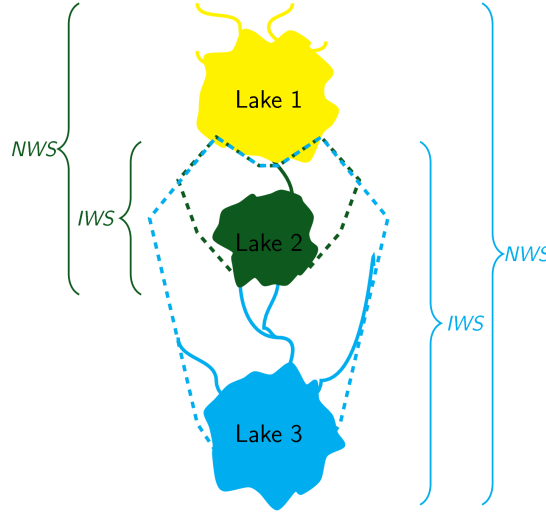

Figure 1: Diagram of the individual lake watershed and network watershed scales. In this example, the lake-watershed for lake 3 includes the lake-watershed of lake 2 because lake 2 is smaller than 10 ha.

We calculated all stream connectivity metrics using the `streamnet` and `nhdR` packages (citations). The algorithms in the `streamnet` package use the `sf` R package as well as the `v.net` and `v.stream.order` modules (Jasiewicz and Metz 2011) included in GRASS GIS (GRASS Development Team 2017). We calculated lake connectivity metrics and non-connectivity lake characteristics using data from the LAGOSNE dataset (Soranno et al. 2017). Both lake and stream connectivity metrics were calculated using the National Hydrography Dataset (NHD) as a primary input. All processed data and code are available at [DOI].

Results

- The relationship between P retention and water residence time was well described by the Vollenweider equations.

- Overall, connectivity played a larger role in determining P retention at the NWS scale than at the IWS scale.

- Lake associated processes played a larger role in determining P retention than stream associated processes at the IWS scale.

- Lakes with both high lake connectivity and high stream connectivity were most often found in the northern portions of the study area.

Table 1: Table of partition splits generated with conditional inference trees. Refer to Figure 1 for scale definitions.

Metric	Split Value	Scale
Baseflow	63.76	iws
Closest lake distance	3773.61	iws
Closest lake distance	2767.04	nws
Average Link Length	2177.08	iws
Average Link Length	2237.34	nws
Max Depth	19.81	misc
Stream order ratio	0.67	iws
Stream order ratio	0.47	nws
Stream density	4.43	iws
Upstream lake area	153.50	nws
Wetland Cover	11.67	iws

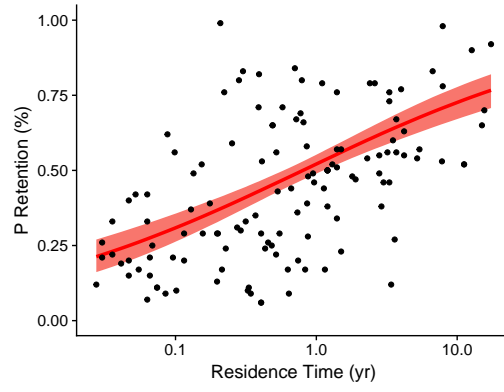**Lake examples?**

Discussion

Figure 2: Population-level fit the Vollenweider P retention model where the dark circles are empirically measured values of P retention and residence time, while the light circles and vertical bars are posterior means and 95% credible intervals respectively.
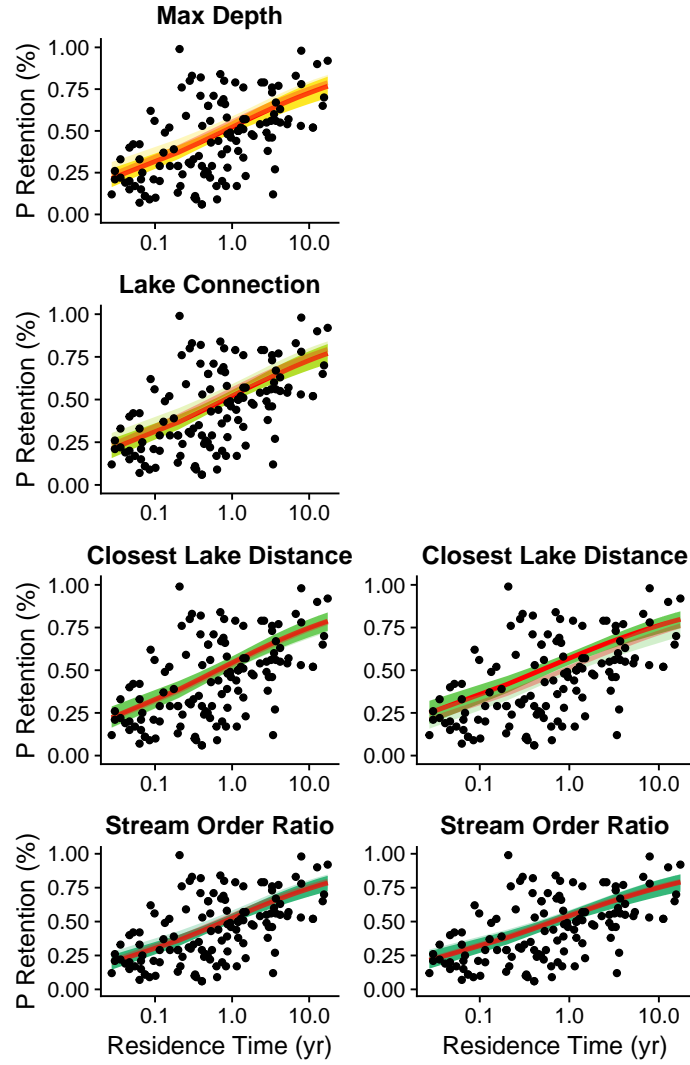
Figure 3: Partitioned (hierarchical) fits of the Vollenweider P retention model where the lines and shaded polygons represent the posterior mean and 95% credible interval respectively. Darker lines and polygons represent the left side of the parition which is usually smaller in magnitude.
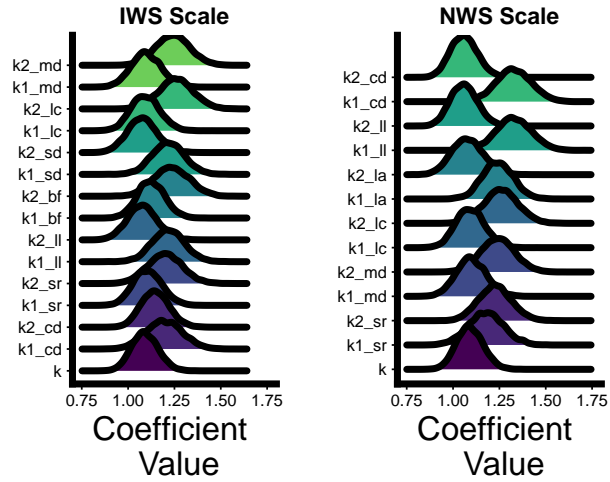
Figure 4: Comparisons of the posterior distribution of Vollenweider's k among lake sub-populations partioned according to Table 1 where the subscript 1 denotes a subpopulation on the left side of the partition which is usually smaller in magnitude.
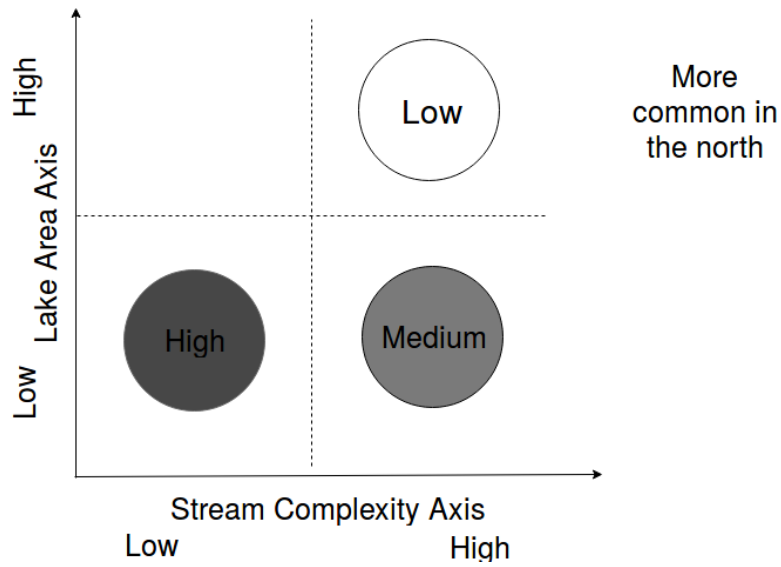

Figure 5: Conceptual diagram