Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes

TYLER WAGNER*, PATRICIA A. SORANNO[†], KATHERINE E. WEBSTER[†,‡] AND KENDRA SPENCE CHERUVELIL[‡,§]

U.S. Geological Survey, Pennsylvania Cooperative Fish & Wildlife Research Unit, Pennsylvania State University, University Park, PA, U.S.A.
[†]*Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, U.S.A.*
[‡]*School of Biological Sciences, Queen's University, Belfast, U.K.*
[§]*Lyman Briggs College, Michigan State University, East Lansing, MI, U.S.A.*

SUMMARY

1. For north temperate lakes, the well-studied empirical relationship between phosphorus (as measured by total phosphorus, TP), the most commonly limiting nutrient and algal biomass (as measured by chlorophyll *a*, CHL) has been found to vary across a wide range of landscape settings. Variation in the parameters of these TP–CHL regressions has been attributed to such lake variables as nitrogen/phosphorus ratios, organic carbon and alkalinity, all of which are strongly related to catchment characteristics (e.g. natural land cover and human land use). Although this suggests that landscape setting can help to explain much of the variation in ecoregional TP–CHL regression parameters, few studies have attempted to quantify relationships at an ecoregional spatial scale.
2. We tested the hypothesis that lake algal biomass and its predicted response to changes in phosphorus are related to both local-scale features (e.g. lake and catchment) and ecoregional-scale features, all of which affect the availability and transport of covarying solutes such as nitrogen, organic carbon and alkalinity. Specifically, we expected that land use and cover, acting at both local and ecoregional scales, would partially explain the spatial pattern in parameters of the TP–CHL regression.
3. We used a multilevel modelling framework and data from 2105 inland lakes spanning 35 ecoregions in six US states to test our hypothesis and identify specific local and ecoregional features that explain spatial heterogeneity in TP–CHL relationships. We include variables such as lake depth, natural land cover (for instance, wetland cover in the catchment of lakes and in the ecoregions) and human land use (for instance, agricultural land use in the catchment of lakes and in the ecoregions).
4. There was substantial heterogeneity in TP–CHL relationships across the 35 ecoregions. At the local scale, CHL was negatively and positively related to lake mean depth and percentage of wooded wetlands in the catchment, respectively. At the ecoregional scale, the slope parameter was positively related to the percentage of pasture in an ecoregion, indicating that CHL tends to respond more rapidly to changes in TP where there are high levels of agricultural pasture than where there is little. The intercept (i.e. the ecoregion-average CHL) was negatively related to the percentage of wooded wetlands in the ecoregion.

Correspondence: Tyler Wagner, U.S. Geological Survey, Pennsylvania Cooperative Fish & Wildlife Research Unit, Pennsylvania State University, 402 Forest Resources Bldg, University Park, PA 16802, U.S.A. E-mail: txw19@psu.edu

5. By explicitly accounting for the hierarchical nature of lake–landscape interactions, we quantified the effects of landscape characteristics on the response of CHL to TP at two spatial scales. We provide new insight into ecoregional drivers of the rate at which algal biomass responds to changes in nutrient concentrations. Our results also indicate that the direction and magnitude of the effects of certain land use and cover characteristics on lake nutrient dynamics may be scale dependent and thus likely to represent different underlying mechanisms regulating lake productivity.

Keywords: Bayesian hierarchical models, ecological drainage unit, eutrophication, lake productivity, total phosphorus–chlorophyll *a* relationships

Introduction

The appeal of the empirical total phosphorus (TP)–chlorophyll *a* (CHL) relationship is demonstrated by the ubiquity of its development for a wide range of lake types and geophysical regions. Studies comparing the properties of published empirical models (e.g. Prairie, Duarte & Kalff, 1989; Brown *et al.*, 2000; Phillips *et al.*, 2008) have shown that the estimated parameters (i.e. the slope and intercept) of the TP–CHL linear regression model vary by lake characteristics (e.g. lake depth) and geographic regions (Malve & Qian, 2006; Jackson *et al.*, 2007; Phillips *et al.*, 2008; Freeman, Lamon & Stow, 2009). These differences in the slope and intercept of the TP–CHL regression have critical management and ecological implications related to the efficiency of nutrient reduction and associated ecosystem responses.

Three factors have been identified as having strong effects on the TP–CHL relationship. First, low N/P ratios generally result in lower CHL at high P concentrations because of N-limitation (Prairie *et al.*, 1989; Downing & McCauley, 1992; McCauley, Downing & Watson, 1998). Second, high concentrations of calcium and associated alkalinity can result in lower CHL per unit P because of chemical immobilisation (Håkanson *et al.*, 2005; Phillips *et al.*, 2008). Third, high humic water colour is typically associated with higher CHL per unit P, although the exact mechanisms have not been identified (Nurnberg & Shaw, 1999; Phillips *et al.*, 2008; Webster *et al.*, 2008). Interestingly, like TP, which typically is supplied to lakes from both natural and human-modified terrestrial sources, all three of these factors have very strong ties to the surrounding landscape. As examples, N/P supply rates from anthropogenic land use activities strongly influence N/P ratios in lakes (Downing &

McCauley, 1992), calcium and alkalinity in lakes reflect geologic setting (Baker *et al.*, 1991), and humic water colour in lakes is derived from surrounding wetlands and forests (Gergel, Turner & Kratz, 1999; Xenopoulos *et al.*, 2003; Canham *et al.*, 2004). These relationships suggest that TP–CHL regressions should have a strong landscape or regional signal, but given the differing mechanisms and spatial heterogeneity of these three factors, the resulting pattern could be complex to quantify or interpret.

Although several studies have quantified ecoregional differences in the TP–CHL parameters (e.g. Phillips *et al.*, 2008), the underlying causes of these differences are less well understood. The primary goal of our paper is to explicitly fit ecoregional TP–CHL models and to explain observed differences in model parameters using ecoregional landscape features, specifically human land use (hereafter land use) and natural land cover (hereafter land cover). Our hypothesis was that lake algal biomass at a given TP, as well as the predicted response rate to changes in phosphorus, is related both to local-scale factors that affect internal nutrient loading, such as lake depth, and to ecoregional-scale landscape features that affect the availability and transport of covarying solutes such as alkalinity, nitrogen and organic carbon. For example, we predicted that land use would influence not only the average CHL concentration of lakes within an ecoregion, but also the average CHL response of lakes within an ecoregion to changes in TP. If this prediction is correct, we would expect correlations between land use characteristics of specific ecoregions and corresponding ecoregion-specific TP–CHL regression parameters.

Ordinary least-squares (OLS) regression is commonly used to estimate parameters for TP–CHL regressions. More recently, however, it has been

demonstrated that Bayesian hierarchical models have many desirable properties, including providing unbiased estimates compared with OLS regression (Malve & Qian, 2006), that make them ideal for estimating nutrient–CHL relationships from data collected across ecoregions or from different lake types (Stow *et al.*, 2009). Further, because aquatic ecosystems are hierarchically organised (Wiens, 1989; Lowe, Likens & Power, 2006), we have to model features at more than one spatial scale. Finally, hierarchical models provide a flexible framework for modelling variation in parameters of cross-sectional data (Gelman & Hill, 2007). Here, we used Bayesian hierarchical models to determine whether land use patterns explain ecoregional parameter variation in the TP–CHL relationship for a diverse set of more than 2000 inland lakes across the midwest and north-eastern US. Because of the complexity inherent in such landscape drivers, our hypothesis can only be addressed using an integrated multilevel modelling framework that can factor in covariates measured at a several spatial scales.

Methods

Lake data

Our database included 2105 lakes in six states: Michigan, Wisconsin, Iowa and Ohio in the Upper midwest, and Maine and New Hampshire in the north-east (Fig. 1). TP, CHL and mean depth data were compiled from the state management agency database described in Webster *et al.* (2008) and data from the Ohio Environmental Protection Agency and Iowa State University. In brief, these data represent single-point measures collected predominantly between 1990 and 2003 (with a few measurements dating back to 1975) from the mixed layer during summer stratification (July to September) from lakes of surface area ≥0.01 km^2 and maximum depth ≥2 m. Each lake is represented by a single data point in the analysis (i.e. there were no repeat observations of individual lakes).

Landscape data

We analysed landscape data at both spatial scales that have an *a priori* relationship with either lake TP or CHL, or the factors that influence TP or CHL: per cent total agriculture, pasture agriculture, arable (row crop) agriculture, total wetlands, herbaceous wetlands, wooded wetlands, and two measures of geologic setting that are related to lake alkalinity: glacial till and patchy quaternary sediment. For land use and cover, we used the 1992 National Land Cover Dataset (http://landcover.usgs.gov/natllandcover.php) as described in Webster *et al.* (2008) and Wagner *et al.* (2007). For the measures of geologic setting, we used the two surficial geology categories in a data set available for the eastern US (Soller & Packard, 1998).

To measure landscape features at 'Level 1' (see Statistical analysis for details), we quantified landscape features around lakes using a 500-m buffer rather than using catchments (because catchment delineations were not available for our 2105 lake data set and were too costly to generate). We used these buffers as a surrogate for catchment land use and cover, recognising that the actual land use and cover in some catchments may differ from those in the 500-m buffer, but in general, these two ways to measure land use and cover will be strongly correlated with

Fig. 1 Map of study area including states (shaded in map and in insert showing location within North America), study lakes (black dots) and the ecoregions (ecological drainage units, thick outlines). One- to three-digit code names are given for each ecoregion in Figs 3 & 4 and full names are in Appendix S1.
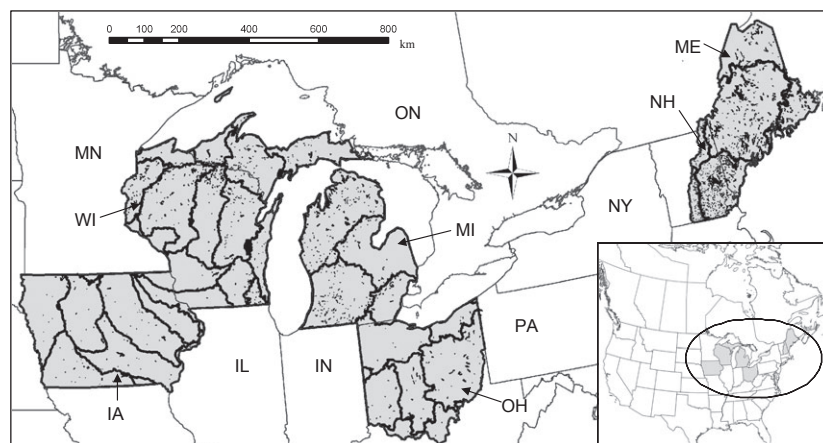
each other. We tested this assumption on a subset of our lakes for which we had both measures (461 MI lakes, P. A. Soranno, unpubl. data). Land use and cover in the 500-m buffer was highly correlated with that in the catchment (% agriculture, $r = 0.79$, % forest, $r = 0.86$ and % wetland, $r = 0.64$). Therefore, the 500-m buffer is a reasonable measure of lake catchment land use and cover.

An additional assumption that we made was regarding the selection of the most appropriate land use and cover dataset. Land use and cover data for our large study area were not available for many years within the time interval the study lakes were sampled, nor were standardised methods used for land use and cover measured at the state level. We chose the 1992 National Land Cover Dataset (http://landcover.usgs. gov/natllandcover.php) because consistent methods were used across the entire study area, and 1992 is close to the median and mean time periods of the limnological sample dates (median year is 1999; mean year is 1995). Ideally, one would exactly match the lake data to the land use and cover data. However, there is little consensus as to whether current or historical land use and cover, both of which can strongly influence fresh waters, represents the more important time period influencing lake nutrients (e.g. Harding *et al.*, 1998; Maloney *et al.*, 2008). Several studies have shown that past land use and cover has very strong, if not stronger, effects on freshwater responses than present land use and cover (e.g., Harding *et al.*, 1998; Maloney *et al.*, 2008). Therefore, the fact that the land use and cover data are older than some of the limnological samples should not be a problem because lake TP probably integrates the effects of past land uses and covers.

To measure landscape features at Level 2, we quantified the same landscape features that we quantified at Level 1 for each of the 35 ecoregions in our study area (see Statistical analysis for details). We chose ecological drainage units (Higgins *et al.*, 2005; Appendix S1) as the regionalisation framework to explore Level 2 variation in TP–CHL regression parameters. Ecological drainage units are based on the agglomeration of river catchments and represent patterns in physiography, climate and freshwater ecosystem connectivity at a coarse spatial scale (Higgins *et al.*, 2005). Ecological drainage units have been shown to outperform other lake grouping schemes in terms of capturing among-region heterogeneity in

water-quality metrics for the state of Michigan (Spence Cheruvelil *et al.*, 2008), and they were identified as an important measure of the freshwater landscape for the six state study area (Spence Cheruvelil, unpubl. data).

Statistical analysis

We used a hierarchical Bayesian modelling framework to identify ecoregional landscape features that explain among-ecoregion variation in TP–CHL regression parameters. The model is a varying intercept, varying slope model and can be viewed as having two levels, each of which can have predictor variables and variance components. The first level of the model has lake-level predictors of CHL, such as lake TP concentration and mean depth, and the second level has ecoregional landscape measurements for modelling variability among the varying intercepts and slopes.

Because our primary goal was to model variation in TP–CHL regression parameters, as an initial modelling step, we quantified the variation in slopes and intercepts in the TP–CHL relationship among regions by fitting a model with $\log_{10}$ CHL as the response variable, $\log_{10}$ TP as the predictor variable and ecoregion as a grouping factor. The linear model was unconditional at Level 2 (i.e. included no landscape predictors). The first level of the model is as follows:

Level 1:

$$y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]} TP_i, \sigma_y^2\right), \text{ for } i = 1, \ldots n \qquad (1)$$

where y_i is $\log_{10}$ CHL from lake i, α_j is the intercept for the jth ecoregion, β_j is the slope for the TP–CHL relationship for the jth ecoregion and σ_y^2 is the residual variance after accounting for lake TP. The second level of the model is as follows:

Level 2:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), \text{ for } j = 1, \ldots J$$

$$(2)$$

where μ_α is the population–average intercept (the average intercept across all ecoregions), μ_β is the population–average slope of the TP–CHL relationship, σ_α^2 and σ_β^2 are the variance estimates among the ecoregion-specific intercepts and among the ecoregion-specific

slopes of the TP–CHL relationships, respectively, and ρ is a between-group correlation parameter. We used non-informative normal priors for μ_α and μ_β and non-informative uniform priors for σ_y, σ_α and σ_β (Gelman, 2006). We modelled the covariance matrix for the intercepts and slopes using the scaled inverse-Wishart distribution (i.e. a uniform prior for ρ; Gelman & Hill, 2007).

After quantifying the variability in slopes and intercepts in the TP–CHL regression parameters among ecoregions, we used the following modelling process to identify the specific landscape predictors that best explained the variation. First, because estimating model parameters within a Bayesian context is computationally intensive, and we were interested in examining the effects of several landscape covariates, we examined the effects of each landscape covariate separately on both the varying intercepts and slopes

by fitting hierarchical linear models using the lmer function of the R package lme4 (Bates & Maechler, 2009; R Development Core Team, 2009). We retained covariates (Table 1) that explained at least some of the variation in the slopes and/or intercepts, as judged by comparing the magnitude of the conditional variance relative to the unconditional variance estimates (i.e. if the conditional variance was smaller than the unconditional variance, we retained the covariate for consideration in future modelling). In addition, because many of the landscape covariates were highly correlated, out of a set of correlated variables, we retained the variable that explained the most variance in the TP–CHL regression parameters. This approach resulted in a set of uncorrelated covariates for use in Bayesian analysis.

We then added Level 2 landscape covariates to the model. The inclusion of covariates to explain variation

Table 1 Summary statistics for the response variable (chlorophyll *a*, CHL) and Level 1 and Level 2 covariates examined to explain within-ecoregion variation in CHL and between-ecoregion variation in TP–CHL regression parameters, respectively

Mechanism		1st quartile	Mean	Median	3rd quartile
NA	Chlorophyll *a* (μg L^{-1})	2.6	10.0	4.4	8.5
	Level 1 covariates				
NA	Total phosphorus (μg L^{-1})[†]	8.0	22.7	12.0	20.0
NA	Mean depth (m)[†]	2.4	4.5	3.7	5.6
Colour	Wetlands, wooded (%)[†]	1.3	7.2	3.5	9.0
Colour	Wetlands, herbaceous (%)	0.3	1.9	1.3	2.6
Colour	Wetlands, total (%)	2.5	8.8	5.0	10.6
N/P	Agriculture, pasture (%)[†]	0.1	5.5	0.8	5.4
N/P	Agriculture, arable (%)	0.6	8.7	3.6	11.3
N/P	Agriculture, total (%)	1.0	14.9	5.5	19.5
Alkalinity	Surficial geology, a^1 (%)	0.0	12.5	0.0	0.0
Alkalinity	Surficial geology, b^2 (%)	0.0	53.5	60.5	100.0
	Level 2 covariates				
Colour	Wetlands, wooded (%)[†]	3.1	6.3	3.3	8.4
Colour	Wetlands, herbaceous (%)	0.9	1.6	1.5	1.7
Colour	Wetlands, total (%)	4.2	7.9	4.8	10.8
N/P	Agriculture, pasture (%)[†]	1.1	6.8	2.0	13.1
N/P	Agriculture, arable (%)	4.5	15.4	5.3	17.7
N/P	Agriculture, total (%)	6.4	22.7	7.2	36.0
Alkalinity	Surficial geology, a^1 (%)	0.2	12.6	14.4	23.3
Alkalinity	Surficial geology, b^2 (%)	47.0	52.9	47.7	65.3

[†]Covariates used in Bayesian hierarchical modelling (see Methods for details).
[1]a is Patchy quaternary sediment.
[2]b is Glacial till.
The ecoregions used in this analysis were ecological drainage units. Total number of lakes (sample size, *n*) was 2105 for Level 1 summaries and 35 ecological drainage units for Level 2 summaries. Percentages of agriculture and wetlands for Level 1 covariates were summarised within an equidistant 500-m buffer around each lake perimeter using GIS. Land use and cover variables for Level 2 covariates were summarised for each ecoregion. Mechanism refers to the hypothesised linkage between the landscape feature and the lake chemical response thought to influence the TP–CHL relationship from the literature; NA refers to variables for which a mechanism is not applicable

in slopes and intercepts, for example the per cent wetlands and agricultural land use within an ecoregion, modifies the Level 2 model in eqn 2 to:

Level 2:

$$\binom{\alpha_j}{\beta_j} \sim N\left(\binom{\gamma_0^\alpha + \gamma_1^\alpha\%\text{wetlands}_j}{\gamma_0^\beta + \gamma_1^\beta\%\text{agriculture}_j}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right),$$
$$\text{for } j = 1,\ldots J \tag{3}$$

where γ_0^α, γ_0^β, γ_1^α and γ_1^β are the coefficients for the intercept, effect of TP and per cent wetlands and agriculture, respectively. σ_α and σ_β are now conditional variances, the regional variance in α_j and β_j after controlling for regional per cent wetlands and agricultural land use. Non-informative normal priors were used for γ_0^α, γ_0^β, γ_1^α and γ_1^β and priors for σ_α, σ_β and ρ are as described in eqn 2. Lastly, because lake-specific properties are important drivers of nutrient dynamics, we considered models that had covariates in addition to TP at Level 1. The Level 1 covariate that we wanted to control for *a priori* was mean depth, because of the importance of lake depth in driving light climate and P dynamics (Vollenweider, 1968). We also examined models that controlled for the Level 1 covariates that were identified as important in explaining Level 2 variation among TP–CHL parameters. We limited the Level 1 covariates to this subset of potentially important Level 1 covariates because our primary goal was to identify land use characteristics that explain variation among parameters. This process of selecting covariates resulted in a set of models that were then fitted and evaluated as described below. Prior to fitting models, we centred all Level 1 covariates by subtracting each value from the overall mean ($x_{ij} - \bar{x}$; i.e. centred on the grand mean) to aid model convergence (Gelman, 2004).

After fitting the set of models, we used the deviance information criterion (DIC; Spiegelhalter *et al.*, 2002) as a measure of fit for the entire model and explained per cent variance for each level (as described in Gelman & Pardoe, 2006) to quantify model fit at each level of the model. We ran three Markov chains in parallel, each with starting values drawn from a random distribution. After discarding the first 5000 iterations of each chain, we examined the scale reduction factor ($\hat{R}$), a convergence statistic, for each parameter. This statistic is the possible reduction in the width of the confidence interval if the simulations

were run forever (Park, Gelman & Kaplan, 2006). We ran simulations until $\hat{R}$ was less than or equal to 1.1 for all parameters (Park *et al.*, 2006). All analyses were performed using WinBUGS version 1.4 (Spiegelhalter *et al.*, 2004). Means are reported ± 1 standard deviation, unless stated otherwise.

Results

Lakes included in our analysis ranged from oligotrophic to eutrophic, with TP ranging from 2 to 765 μg L^{-1} and CHL ranging from 0.1 to 328 μg L^{-1} (Table 1). Mean TP and CHL, across all lakes, were 22.7 ± 41.8 and 10.0 ± 20.5 μg L^{-1}, respectively. Lake mean depth and landscape characteristics surrounding individual lakes and among ecoregions also varied considerably. For example, percentage pasture surrounding individual lakes ranged from 0 to 80% and percentage pasture within an ecoregion ranged from 0.3 to 39% (Table 1).

Our focus on the relationship between TP and CHL relies on the assumption that P is the primary factor limiting algal biomass in the lakes. Because N-limitation has been invoked in other studies, we investigated the relative importance of N – compared with P – limitation by examining the conditional relationship between TP, total nitrogen (TN) and CHL (Appendix S2; Qian, 2010). The conditional plots show the relationship between TP–CHL at relatively constant TN concentrations. The same was done for TN–CHL relationship holding TP constant. These analyses suggested that P, rather than N, was most important in limiting algal biomass in our study lakes.

TP–CHL relationship among regions

As expected, CHL was positively correlated with TP, the latter explaining 56% of the variability in CHL (Fig. 2). The estimated posterior means for the population–average intercept and slope were 0.70 [95% credible interval (CI) = 0.64, 0.76] and 0.86 (95% CI = 0.77, 0.97), respectively (Table 2). However, there was substantial variation among ecoregions in both the intercepts and slope parameters of the TP–CHL relationship (Table 2; Fig. 2). Estimated posterior means for variation among intercepts (σ_α) and slopes (σ_β) was 0.14 (95% CI = 0.10, 0.20) and 0.21 (95% CI = 0.12, 0.33), respectively. Estimated posterior

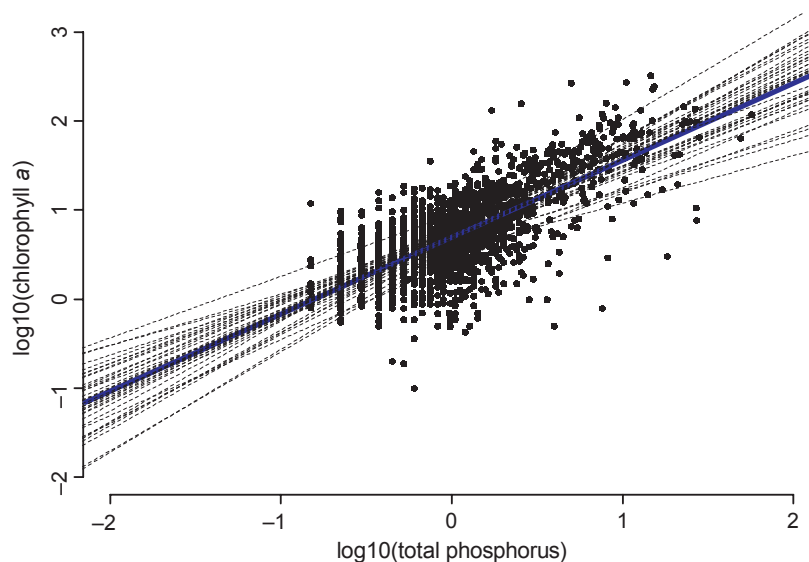**Fig. 2** Relationship between $\log_{10}$ total phosphorus (μg L^{-1}; grand-mean centred, see Methods) and $\log_{10}$ chlorophyll *a* (μg L^{-1}). Solid line is the estimated population–average relationship and dashed lines are ecoregion-specific regression relationships from a Bayesian hierarchical model (eqn 1).

means for ecoregion-specific intercepts ranged from 0.40 to 0.95 and for slopes ranged from 0.53 to 1.15 (Figs 3 & 4). Because we centred on the grand mean of TP, the intercept for the *j*th ecoregion is interpreted as the predicted CHL concentration ($\log_{10}$-transformed) in a lake whose $\log_{10}$-transformed TP value is equal to the grand mean (i.e. an adjusted mean for the *j*th ecoregion). Ecoregions in Iowa, southern Wisconsin and eastern Ohio had, on average, the highest intercept estimates (i.e. higher average CHL concentrations) compared with other ecoregions (Fig. 3). Ecoregions in central Wisconsin and parts of Iowa and northern Ohio also had, on average, higher slope estimates compared with other ecoregions. In contrast, regions in northern Michigan and Maine and New Hampshire tended to have the lowest estimates of slope and intercept (Fig. 4).

Model selection: explaining variation among ecoregion-specific TP–CHL regression parameters

The Level 2 covariates (to explain variation in parameters among ecoregions) included percentage pasture agriculture for explaining variation among slopes and percentage wooded wetlands for explaining variation among intercepts. Because percentage wooded wetlands data were skewed towards zero, a logit transformation was used. Using these covariates, we fitted a total of six candidate models (Table 2). Using DIC, the addition of Level 2 covariates to explain variation among slopes and intercepts pro

vided a superior model than one not including ecoregional predictor variables (DIC = -118.6 for unconditional Level 2 model and -3736.8 for conditional Level 2 model and a smaller DIC value indicates a better fitting model; Table 2). Therefore, all subsequent candidate models included the percentages of pasture and wooded wetlands to model variation among slopes and intercepts, respectively (i.e. models were conditional at Level 2). As a result, the candidate models only differ in complexity by the number and combination of covariates included at Level 1.

Based on DIC, model 5 was the top-ranked model. Covariates at Level 1 were TP (positive correlation), mean depth (negative correlation) and percentage wetlands (positively correlation); at Level 2, percentage pasture was the covariate for the slopes and percentage wooded wetlands the covariate for the intercepts (Table 2). The model explained 58% of the variation in CHL. Percentage wooded wetlands explained 42% of the variation among ecoregion-specific intercepts, with ecoregion-average CHL decreasing with increasing wooded wetlands (Fig. 5a). Percentage pasture explained 37% of the variation among ecoregion-specific slopes, with ecoregion-specific TP–CHL regression slopes increasing with increasing levels of ecoregion pasture (Fig. 5b; Table 2). Although not contained in the top-ranked model, it is worth noting that, at Level 1, CHL was also positively correlated with the percentage of pasture surrounding each lake.

Table 2 Deviance information criterion (DIC) and posterior means followed by 95% credible intervals in parentheses for model parameters

Model	DIC	Level 1 covariates					Level 2 covariates		Standard deviations			Variance explained			
		Intercept (γ_0^α)	TP (γ_0^β)	Depth (δ_1)	Ag (δ_2)	Wetland (δ_3)	Wetland (γ_1^β)	Ag (γ_1^β)	σ_y	σ_α	σ_β	r_y^2	r_α^2	r_β^2	ρ
1	–118.60	0.70 (0.64, 0.76)	0.86 (0.77, 0.97)	–	–	–	–	–	0.29 (0.28, 0.30)	0.14 (0.10, 0.20)	0.21 (0.12, 0.33)	0.56	–	–	–0.04
2	–3736.80	0.52 (0.41, 0.64)	0.69 (0.56, 0.82)	–	–	–	–0.05 (–0.08, –0.02)	1.19 (0.43, 2.03)	0.29 (0.28, 0.30)	0.10 (0.06, 0.16)	0.15 (0.02, 0.28)	0.57	0.33	0.30	–0.03
3	–2612.30	0.52 (0.41, 0.63)	0.64 (0.52, 0.77)	–0.009 (–0.01, –0.005)	–	–	–0.05 (–0.08, –0.02)	1.30 (0.54, 2.13)	0.29 (0.28, 0.30)	0.10 (0.06, 0.15)	0.16 (0.05, 0.27)	0.57	0.34	0.34	–0.08
4	–9497.50	0.54 (0.42, 0.66)	0.6 (0.51, 0.76)	–0.01 (–0.015, –0.006)	0.03 (0.02, 0.04)	–	–0.04 (–0.08, –0.007)	1.21 (0.48, 2.03)	0.29 (0.28, 0.30)	0.12 (0.08, 0.17)	0.14 (0.01, 0.27)	0.58	0.22	0.32	–0.14
5	–21 737.20	0.48 (0.36, 0.60)	0.65 (0.53, 0.75)	–0.009 (–0.01, –0.004)	–	0.04 (0.02, 0.06)	–0.06 (–0.10, –0.03)	1.29 (0.64, 2.06)	0.29 (0.28, 0.30)	0.11 (0.07, 0.16)	0.11 (0.006, 0.25)	0.58	0.42	0.37	0.02
6	–10 427.60	0.49 (0.37, 0.62)	0.63 (0.50, 0.76)	–0.01 (–0.02, –0.005)	0.03 (0.01, 0.04)	0.04 (0.02, 0.06)	–0.06 (–0.09, –0.02)	1.28 (0.51, 2.09)	0.29 (0.28, 0.30)	0.12 (0.08, 0.18)	0.15 (0.02, 0.28)	0.58	0.30	0.33	–0.20

TP, total phosphorus; Ag, % pasture; wetland, % wooded wetland; depth, mean depth. The effects of Ag, wetland and depth are designated by δ_i. See Table 1 for units and summary statistics for covariates. For Level 2 covariates, per cent pasture land use and per cent wooded wetlands are modelling variation among slopes and intercepts in the TP–chlorophyll *a* relationships, respectively. r_y^2, r_α^2 and r_β^2 are the per cent variance explained within-regions, among intercepts and among slopes, respectively. See Methods eqns 1–3 for model details.

Discussion

We found substantial ecoregional variation in TP–CHL relationships across the 35 ecoregions, which was manifest as important differences in both intercepts and slopes. More importantly, we identified relationships between these intercepts and slopes and ecoregional landscape features that included two important classes of land use and cover: the percentages of pasture and wooded wetlands. Our results help to build upon previous studies that examined only the local-scale factors influencing the TP–CHL relationship. By explicitly considering the surrounding landscape at two spatial scales (the lake catchment and the overall ecoregional context), as well as the in-lake features that these landscape drivers influence, we have a more complete understanding of factors that control within – and across – ecoregion dynamics of the TP–CHL relationship. Further, our results support the hypothesis that lake algal biomass at a given TP, as well as the average predicted response rate to changes in phosphorus, is related to both local-scale and ecoregional-scale landscape features. For example, we predicted that land use and cover would influence not only the average CHL concentration of lakes within an ecoregion, but also the average CHL response of lakes in that ecoregion to changes in TP. Thus, our prediction that land use characteristics of specific ecoregions were correlated with corresponding ecoregion-specific TP–CHL regression parameters was confirmed. We offer a synthetic depiction of these linkages in a flow diagram that shows the landscape and lake drivers of the TP–CHL relationship (Fig. 6).

The ranges of estimated intercepts and slopes from the TP–CHL hierarchical model were similar to others reported in the literature. Our estimated ecoregion-specific slopes ranged from 0.53 to 1.15, with a population–average mean slope of 0.86. Estimated slopes of TP–CHL OLS regressions reported in Phillips *et al.* (2008), and references therein, ranged from 0.72 to 1.4, with a mean slope of 0.94 (*n* = 15). Because we grand-mean-centred log$_{10}$(TP), our intercept estimates were positive as opposed to the negative intercepts commonly reported in the literature for non-centred analyses and thus are not directly comparable except in a relative sense.

Previous studies have focused primarily on examining the ability of land use and cover characteristics
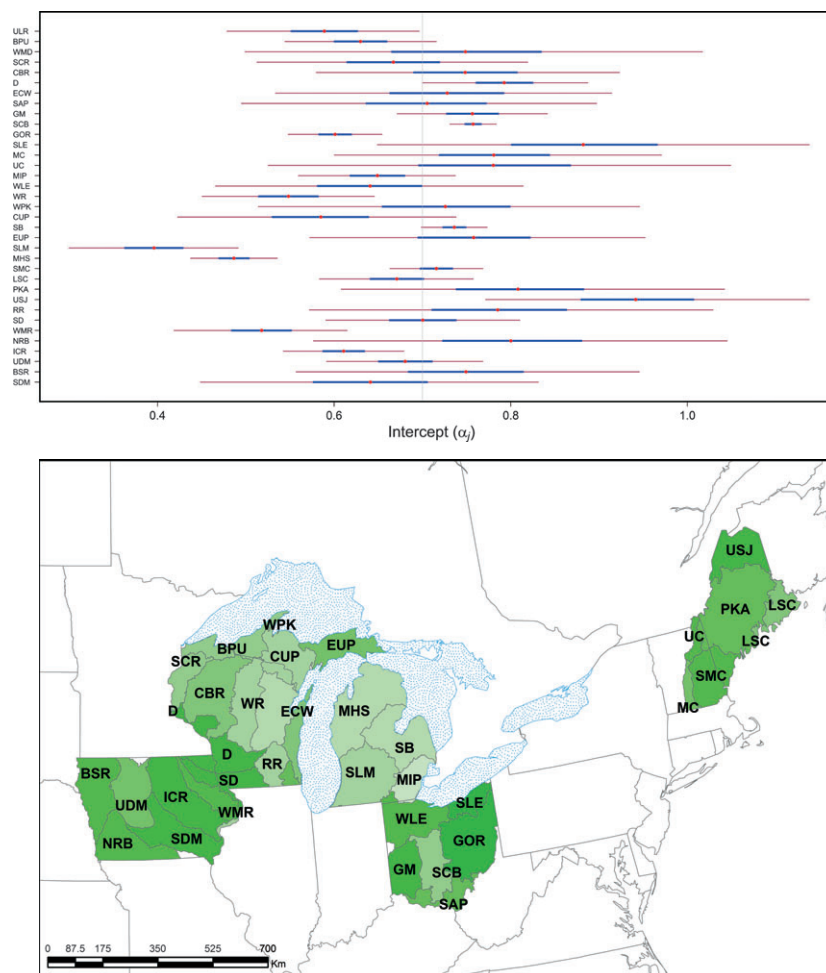
Fig. 3 Region-specific intercepts (α_j) for the total phosphorus–chlorophyll *a* regression. Solid points are the estimated posterior means, thick lines are 50% posterior CIs, and thin lines are 95% CIs. Vertical line represents population–average posterior mean estimate. Three-digit codes on the y-axis correspond to ecoregions (see Appendix S1). The map depicts variation in intercepts among ecoregions, where darker shading indicates larger intercepts.

quantified at the local scale (i.e. Level 1 in our model) to explain variation in water-quality metrics among waterbodies. In our analysis, the top-ranked model contained a positive relationship between the percentage of wooded wetlands and CHL; however, we also observed a positive correlation between CHL concentrations and percentage pasture in several candidate models. The positive relationship between lake productivity and agricultural land use in the lake catchment is expected and has been well established (Carpenter *et al.*, 1998). The positive relationship between CHL and the percentage of wooded wetlands could be because of the influence of humic colour, supplied to lakes from wetlands, on autotrophic production. Webster *et al.* (2008), using a subset of the database we used in this analysis, explored interactions between TP and humic colour and effects of the two on CHL. They summarised several potential mechanisms that could be responsible for

higher CHL concentrations in humic lakes, such as phytoplankton communities being restricted to a smaller volume of water because of a reduction in the photic zone. In addition to TP and percentage wooded wetlands, our top-ranked model contained mean depth as a Level 1 predictor negatively correlated with CHL. This negative correlation was expected, because lake mean depth has been shown to be a major influence on P dynamics in lakes, largely because of the importance of internal nutrient loading from the sediments in shallow lakes (Wetzel, 2001; Søndergaard, Jensen & Jeppesen, 2003). Therefore, our Level 1 model substantiates past research establishing relationships between CHL and land use or cover and mean depth.

In contrast to the positive relationship between percentage wooded wetlands in the catchment and CHL, we observed a negative relationship between percentage wooded wetlands in the ecoregion and the
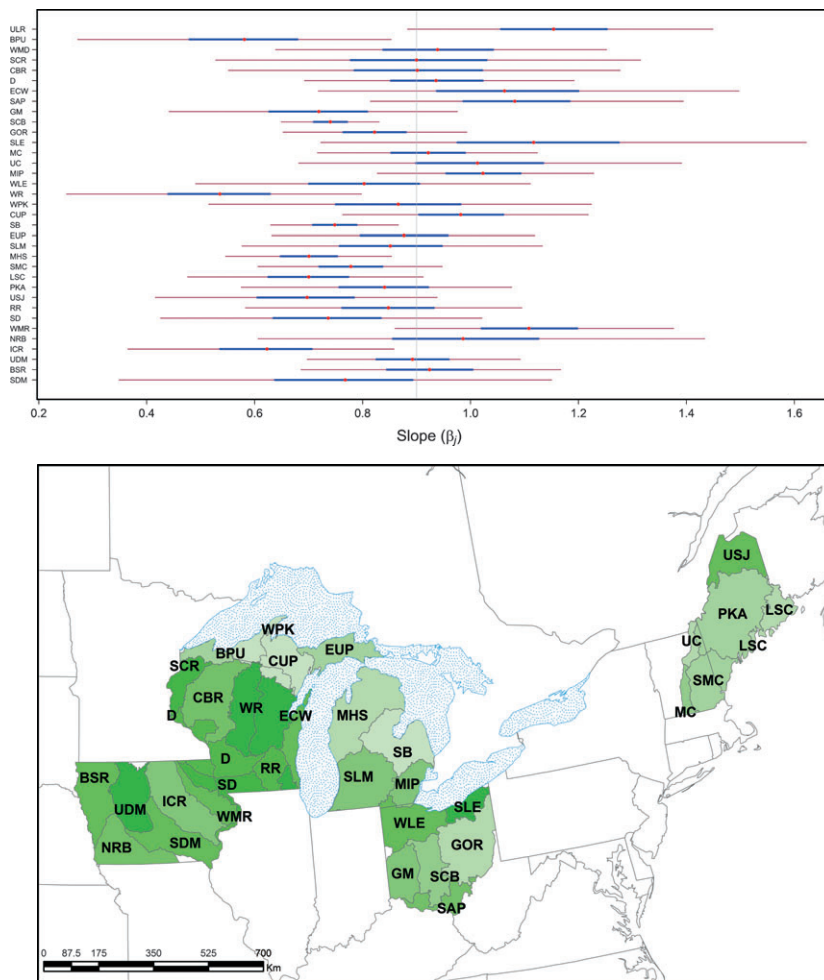
Fig. 4 Region-specific slopes $\left(\beta_j\right)$ for the total phosphorus–chlorophyll *a* regression. Solid points are the estimated posterior means, thick lines are 50% posterior CIs, and thin lines are 95% CIs. Vertical line represents population–average posterior mean estimate. Three-digit codes on the y-axis correspond to ecoregions (see Appendix S1). The map depicts variation in intercepts among ecoregions, where darker shading indicates larger slopes.

ecoregion-specific intercepts of the TP–CHL regression. This result suggests that, on average, lakes within ecoregions with high wetland land cover have lower CHL concentrations. This contradiction between relationships quantified at different spatial scales [compared with the positive relationship between wetlands and CHL when quantified at the lake-scale (i.e. Level 1)] illustrates a very important contribution of our work: the spatial scale-dependence in the direction of the effect of landscape-based influences on lake productivity. The fact that lakes within an ecoregion with high per cent wetlands have, on average, lower CHL concentrations corresponds to relationships between land use and cover and lake productivity. As natural land cover types such as wetlands are lost and replaced with urban and/or agricultural land uses, we would expect nutrients (and therefore CHL) to increase. However, this pattern can also be influenced by a covariation

between soil productivity and per cent wetlands. For instance, lakes in ecoregions with a high percentage of wooded wetlands may also be those ecoregions with relatively low nutrients, and therefore, lakes in those ecoregions are 'naturally' less productive.

In addition to the negative relationship between percentage wooded wetlands and ecoregion-specific intercepts, we also observed a positive relationship between percentage pasture in the ecoregions and the slope of the TP–CHL relationship. We know that runoff from agricultural activities is a source of nutrients (including N and P) that contributes to the eutrophication of aquatic ecosystems throughout the world (Turner & Rabalais, 2004; Renwick *et al.*, 2008). In a meta-analysis, Taranu & Gregory-Eaves (2008) found that the percentage of land under agricultural use in a catchment explained 28% of the variation in lake TP concentration. Although a correlation does not indicate causation, our results suggest not only
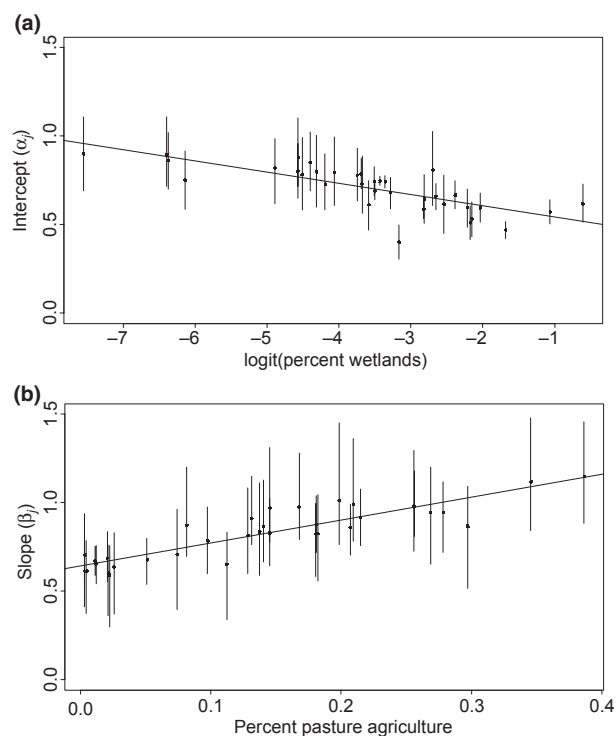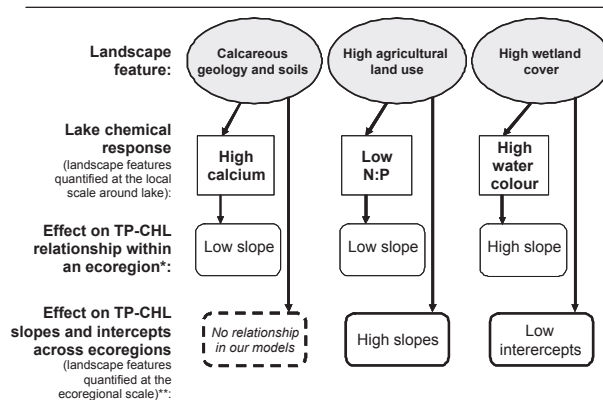
Fig. 5 Ecoregion-specific intercepts (α_j) versus ecoregion per cent wooded wetlands (logit-transformed) (a) and ecoregion-specific slopes (β_j) versus ecoregion per cent pasture agriculture (b) from the total phosphorus–chlorophyll *a* Bayesian hierarchical model (Model 3 in Table 2). Points are estimated posterior means, and vertical lines are 95% posterior credible intervals. Solid line is estimated hierarchical regression line.

Fig. 6 Schematic that is a synthesis of the literature and results from this study illustrating the important landscape influences on the lake chemical responses that affect within-ecoregion and across-ecoregion total phosphorus–chlorophyll *a* regression parameters. Direction of arrows within boxes indicates the direction of the effect of the feature in the oval above the box.

that pasture in a catchment potentially affects lake nutrient *concentrations* and therefore lake productivity (i.e. CHL concentrations), but also that it can influence the *rate* at which CHL, on average, responds to changes in TP. For example, the steeper slope in the TP–CHL regression for lakes located in ecoregions with a high percentage of pasture indicates that CHL potentially increases at a faster rate in response to changes in TP compared with lakes in ecoregions with low percentages of pasture.

We further examined the land use type 'pasture agriculture' in order to understand this contribution of our research more fully. This land use category includes a potentially wide variety of agricultural practices. For instance, the National Land Cover Dataset defines 'pasture agriculture' as 'Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.' Because this category includes land that

contains animals on the land, or land that produces crops for feeding animals elsewhere, nutrients can be exported to aquatic systems through several pathways, including direct fertiliser applications, manure, or vegetation and soil reservoirs (McDowell, Nash & Robertson, 2007; Dougherty *et al.*, 2008). The potential direct and indirect effects of such agricultural activities on aquatic systems, combined with the large spatial scale of this study, do not allow for the elucidation of mechanisms responsible for the observed patterns. However, we speculate that the positive correlation observed between regional pasture and TP–CHL slopes may be because of simultaneous N and P enrichment, reflected in a greater CHL response to TP, when compared with regions dominated by other land uses or covers. Although freshwater ecosystems have historically been viewed as P limited, recent work suggests that co-limitation by both N and P is most common. Elser *et al.* (2007) suggest that a potential mechanism for this co-limitation is that enrichment with N or P quickly results in limitation by the other. Alternatively, areas of intense agriculture could result in lakes with high export of both N and P, as well as high N/P ratios, on average, because of high N inputs relative to P. This could result in lakes that are P limited and thus would respond to changes in P more rapidly than lakes with lower N/P ratios. To examine this hypothesis in more
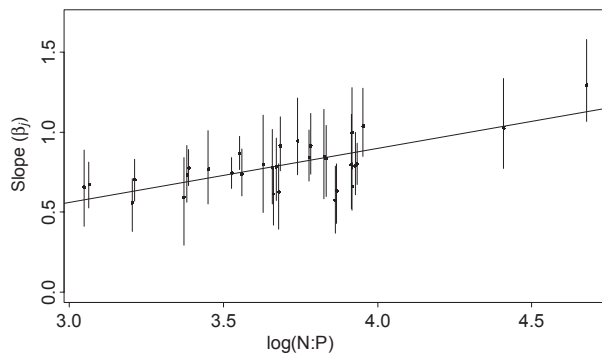
Fig. 7 Ecoregion-specific slopes (β_j) versus ecoregional average N⁄P ratio from the total phosphorus–chlorophyll *a* Bayesian hierarchical model. Points are estimated posterior means, and vertical lines are 95% posterior credible intervals. Solid line is estimated hierarchical regression line.

detail, we selected the subset of our study lakes for which we had N data (n = 1364 lakes and 32 ecological drainage units) and fitted a hierarchical model that included average N⁄P ratio as a predictor of ecoregion-specific TP–CHL slopes. This analysis suggested that ecoregions with higher average N⁄P ratios did in fact have greater TP–CHL slope parameter estimates (Fig. 7), which supports studies showing the influence of N⁄P ratio on TP–CHL relationships (e.g. Downing & McCauley, 1992). However, some research has shown that pasture can export nutrients at low N⁄P ratios when animals are present (Arbuckle & Downing, 2001), which suggests that, in our study area, animals are not present in large numbers on pasture (i.e. the N and P 'signals' at the ecoregional scale may reflect mainly non-animal contributions, such as fertiliser applications). However, because the National Land Cover Dataset definition of pasture agriculture includes such a wide variety of agricultural practices, it is not possible to make inferences on the relative importance of animal versus other sources of pasture-related N and P sources.

Our results suggest that the direction and magnitude of the effects of land use practices on lake nutrient dynamics may be scale-dependent. In addition, these practices may influence not only the concentrations of limiting nutrients, but also the rate at which aquatic ecosystems respond to natural and anthropogenic disturbances. We need additional research conducted across spatial scales to elucidate dynamics, including response rates, related to other important freshwater processes.

References

Arbuckle K.E. & Downing J.A. (2001) Use on lake N:P in a predominantly agricultural landscape. *Limnology and Oceanography*, **46**, 970–975.

Baker L.A., Herlihy A.T., Kaufmann P.R. & Eilers J.M. (1991) Acidic lakes and streams in the United States: the role of acidic deposition. *Science*, **252**, 1151–1154.

Bates D. & Maechler M. (2009) lme4: linear mixed-effects models using S4 classes. R package version 0.999375-30. http://cran.r-project.org/.

Brown C.D., Hoyer M.V., Bachmann R.W. & Canfield D.E. Jr (2000) Nutrient-chlorophyll relationships; an evaluation of empirical nutrient-chlorophyll models using Florida and north-temperate lake data. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 1574–1583.

Canham C.D., Pace M.L., Papaik M.J., Primack A.G.B., Roy K.M., Maranger R.J. *et al.* (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 A.N., Sharpley A.N. & Smith V.H. (1998) Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, **8**, 559–568.

Dougherty W.J., Nicholls P.J., Milham P.J., Havilah E.J. & Lawrie R.A. (2008) Phosphorus fertilizer and grazing management effects on phosphorus in runoff from dairy pastures. *Journal of Environmental Quality*, **37**, 417–428.

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

Elser J.J., Bracken M.E.S., Cleland E.E., Gruner D.S., Harpole W.S., Hillebrand H. *et al.* (2007) Global anal-

ysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial eco-systems. *Ecology Letters*, **10**, 1135–1142.

Freeman A.M., Lamon E.C. III & Stow C.A. (2009) Nutrient criteria for lakes, ponds, and reservoirs: a Bayesian TREED model. *Ecological Modelling*, **220**, 630–639.

Gelman A. (2004) Parameterization and Bayesian modeling. *Journal of the American Statistical Association*, **99**, 537–545.

Gelman A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **3**, 515–534.

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

Gelman A. & Pardoe I. (2006) Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*, **48**, 241–251.

Gergel S.E., Turner M.G. & Kratz T.K. (1999) Dissolved organic carbon as an indicator of the scale of watershed influence on lakes and rivers. *Ecological Applications* **9**, 1377–1390.

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

Harding J.S., Benfield E.F., Bolstad P.V., Helfman G.S. & Jones B.D. (1998) Stream biodiversity: the ghost of land use past. *Proceedings of the National Academy of Sciences U.S.A.*, **95**, 14843–14847.

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.

Jackson L.J., Lauridsen T.L., SØndergaard M. & Jeppesen E. (2007) A comparison of shallow Danish and Canadian lakes and implications of climate change. *Freshwater Biology*, **52**, 1782–1792.

Lowe W.H., Likens G.E. & Power M.E. (2006) Linking scales in stream ecology. *BioScience*, **56**, 591–597.

Maloney K.O., Feminella J.W., Mitchell R.M. & Miller S.A. (2008) Landuse legacies and small streams: identifying relationships between historical land use and contemporary stream conditions. *Journal of the North American Benthological Society*, **27**, 280–294.

Malve O. & Qian S.S. (2006) Estimating nutrients and chlorophyll a relationships in Finnish lakes. *Environmental Science and Technology*, **40**, 7848–7853.

McCauley E., Downing J.A. & Watson S. (1998) Sigmoid relationships between nutrients and chlorophyll among lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 1171–1175.

McDowell R.W., Nash D.M. & Robertson F. (2007) Sources of phosphorus lost from a grazed pasture receiving simulated rainfall. *Journal of Environmental Quality*, **36**, 1281–1288.

Nurnberg G.K. & Shaw M. (1999) Productivity of clear and humic lakes: nutrients, phytoplankton, bacteria. *Hydrobiologia*, **382**, 97–112.

Park D.K., Gelman A. & Kaplan N. (2006) R2WinBUGS: running WinBUGS from R. *The Political Methodologist*, **14**, 5–10.

Phillips G., Pietiläinen O.-P., Carvalho L., Solimini A., Lyche Solheim A. & 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.

Qian S.S. (2010) *Environmental and Ecological Statistics with R*. Chapman & Hall CRC Press, Boca Raton, FL.

R Development Core Team. (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Renwick W.H., Vanni M.J., Zhang Q. & Patton J. (2008) Water quality trends and changing agricultural practices in a Midwest U.S. watershed, 1994 – 2006. *Journal of Environmental Quality*, **37**, 1862–1874.

Soller D.R. & Packard P.H. (1998) Digital representation of a map showing the thickness and character of quaternary sediments in the glaciated United States east of the Rocky Mountains. U.S. Geological Survey Digital Data Series DDS-38.

Søndergaard M., Jensen J.P. & Jeppesen E. (2003) Role of sediment and internal loading of phosphorus in shallow lakes. *Hydrobioogia*, **506–509**, 135–145.

Spence Cheruvelil K., Soranno P.A., Bremigan M.T., Wagner T. & Martin S. (2008) Grouping lakes for water quality assessment and monitoring: the roles of regionalization and spatial scale. *Environmental Management*, **41**, 425–440.

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

Spiegelhalter D.J., Thomas A., Best N. & Lunn D. (2004) WinBUGS User Manual. Version 1.4.

Stow C.A., Lamon E.C., Quian S.S., Soranno P.A. & Reckhow K.H. (2009) Bayesian hierarchical/multilevel models for inference and prediction using cross-system lake data. In: *Real World Ecology: Large-Scale and Long-Term Case Studies and Methods* (Eds S. Miao, S. Carstenn & M. Nungesser), pp. 111–136. Springer Verlag, New York.

Taranu Z.E. & Gregory-Eaves I. (2008) Quantifying relationships among phosphorus, agriculture, and lake depth at an inter-regional scale. *Ecosystems*, **11**, 715–725.

Turner R.E. & Rabalais N.N. (2004) Suspended sediment, C, N, P, and Si yields from the Mississippi River Basin. *Hydrobiologia*, **511**, 79–89.

Vollenweider R.A. (1968) Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication. Paris: OECD Technical Report. p. 159.

Wagner T., Bremigan M.T., Spence Cheruvelil K., Soranno P.A., Nate N.N. & Breck J.E. (2007) A multilevel modeling approach to assessing regional and local landscape features for lake classification and assessment of fish growth rates. *Environmental Monitoring and Assessment*, **130**, 437–454.

Webster K.E., Soranno P.A., Spence Cheruvelil K., Bremigan M.T., Downing J.A., Vaux P.D. et al. (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*, 3rd edn. Academic Press, San Diego, CA, pp. 245–246.

Wiens J.A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

Xenopoulos M.A., Lodge D.M., Frentress J., Kreps T.A., Bridgham S.D., Grossman E. et al. (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.

Supporting Information

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

Appendix S1. Ecological Drainage Unit codes and names (Higgins *et al.*, 2005).

Appendix S2. Conditional plots for log chlorophyll *a* (CHL) versus log total phosphorus (TP) conditional on log total nitrogen (TN) (a) and CHL versus TN conditional on TP (b). TP and TN were centred prior to analysis (see Methods).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copyedited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

(Manuscript accepted 27 April 2011)