Article

# Assessing the influence of watershed characteristics on chlorophyll *a* in waterbodies at global and regional scales

Whitney M. Woelmer,[1] Yu-Chun Kao,[1]* David B. Bunnell,[2] Andrew M. Deines,[3] David H. Bennion,[4] Mark W. Rogers,[2,a] Colin N. Brooks,[5] Michael J. Sayers,[5] David M. Banach,[5] Amanda G. Grimm,[5] and Robert A. Shuchman[5]

[1] *Center for Systems Integration and Sustainability, Michigan State University Department of Fisheries and Wildlife, East Lansing, MI, USA, under contract to US Geological Survey Great Lakes Science Center*
[2] *US Geological Survey Great Lakes Science Center, Ann Arbor, MI, USA*
[3] *Exponent, Inc. Bellevue, WA, USA*
[4] *CSS-Dynamac, Fairfax, VA, USA, under contract to US Geological Survey Great Lakes Science Center*
[5] *Michigan Tech Research Institute, Ann Arbor, MI, USA*
[a] *Current address: US Geological Survey, Tennessee Cooperative Fishery Research Unit, Tennessee Technological University, Cookeville, TN, USA*
* *Corresponding author: kaoyc@msu.edu*

## Abstract

Predictions of chlorophyll *a* (Chl-*a*) in lentic waterbodies (lakes and reservoirs) are valuable to researchers and resource managers alike but have been rarely conducted at the global scale. With the development of remote sensing technologies, it is now feasible to gather large amounts of data across the world, including understudied and remote regions. To determine which factors were most important in explaining the variation of Chl-*a* in waterbodies at global and regional scales, we first developed a database of 227 globally distributed waterbodies and watersheds with corresponding Chl-*a*, nutrient, hydrogeomorphic, and climate data. Then we used a generalized additive modeling approach and selected models that most parsimoniously related Chl-*a* to predictor variables for all 227 waterbodies and for a subset of 51 within the Laurentian Great Lakes region. Our best global model contained 3 hydrogeomorphic variables (waterbody area, shoreline development index, and watershed to waterbody area ratio) and a climate variable (mean temperature in the warmest quarter) that explained about 30% of variation in Chl-*a*. Our regional model contained one hydrogeomorphic variable (watershed area), the same climate variable, and a nutrient variable (percent of watershed area cover by waterbodies) that explained 58% of variation in Chl-*a*. Our results indicate that a regional approach to watershed modeling may be more informative to predicting Chl-*a* than a global approach and that nearly a third of global variation in Chl-*a* may be explained using hydrogeomorphic and climate variables.

**Key words:** chlorophyll *a*, climate, global scale, hydrogeomorphic characteristics, lakes, nutrients, regional scale, reservoirs

## Introduction

Freshwater lentic waterbodies (i.e., lakes and reservoirs; hereafter, waterbodies) provide multiple ecosystem services including transportation, food production, and drinking water (Postel and Carpenter 2012). Many of these services are dependent on the level of primary production. Increases in primary production may lead to increases in the production of lower trophic levels (i.e., zooplankton and benthic invertebrates) and ultimately result in desirable increases in fisheries harvests (Oksanen et al. 1981). Excessive primary production, however, may result in hypoxia, shifts in thermal and light regimes, and other water quality problems that negatively influence wildlife habitat and human drinking water quality (Smith 2003, Smith and Schindler 2009). A growing body of literature in

landscape limnology seeks to understand the extent to which watershed characteristics influence primary production (e.g., D'Arcy and Carignan 1997, Knoll et al. 2003, Nõges 2009, Miranda et al. 2014). Although internal processes such as nutrient cycling or the composition and abundance of the herbivore community (e.g., Vanni 2002, Gruner et al. 2008) undoubtedly influence primary production, watershed characteristics such as land cover may be more easily manipulated by managers than internal influences. In building models to elucidate landscape-level effects on primary production, many scientists use chlorophyll *a* (Chl-*a*) as a surrogate (e.g., Cottingham and Carpenter 1998) because it is easier to measure. Although Chl-*a* is a measure of phytoplankton biomass rather than primary production, meta-analyses reveal Chl-*a* to be generally positively correlated with primary production (Brylinsky and Mann 1973, Morin et al. 1999), although some variation remains unexplained. Models that can predict Chl-*a* from remotely measured aspects of the watershed or climate would be particularly useful in situations where the number of waterbodies exceeds the capacity of scientists and managers to conduct direct measurements.

With few exceptions (e.g., Brylinsky and Mann 1973, Meeuwig and Peters 1996), most studies that have sought to link watershed characteristics to production of downstream waterbodies have been limited to specific regions at relatively small spatial scales. A majority of these regional studies have revealed the importance of watershed land cover to Chl-*a* in waterbodies (e.g., Chen and Lei 2001, Knoll et al. 2003, 2015, Jones et al. 2004, Miranda et al. 2014, Huo et al. 2015). Often, the percent of watershed area dominated by agriculture has the strongest positive correlation to Chl-*a* (Carpenter et al. 1998, Jones et al. 2004), but this finding is not universal, with many studies finding that other types of land cover, such as urban or forest, better explain Chl-*a* in different regions (Liu et al. 2011, Huo et al. 2015, Knoll et al. 2015) or that land cover fails to explain the variation in Chl-*a* (Burford et al. 2007, Carneiro et al 2014). In addition, a more globally distributed study by Meeuwig and Peters (1996) showed that forest cover and human population in the watershed were important predictor variables to explain Chl-*a* variation among 63 lakes.

Hydrogeomorphic (HGM) characteristics of the watershed or waterbody also can be important predictors of Chl-*a* in regional studies. For example, catchment slope and area accounted for up to 60% of the variation in Chl-*a* in 32 Canadian Shield lakes (D'Arcy and Carignan 1997), whereas Liu et al. (2011) and Martin et al. (2011) found that mean depth was a more important predictor of Chl-*a* in 150 Michigan lakes and 19 lakes in the Yunnan plateau and lower Yangtze floodplain in China, respectively. Other

studies found that the ratio of watershed to waterbody area best predicted physicochemical parameters related to Chl-*a*, for example in Swedish lakes and Ohio reservoirs (Knoll et al. 2003, Håkanson 2005, Bremigan et al. 2008). Additionally, in a study including 1337 lakes across Europe, Nõges (2009) reported depth to be inversely correlated with Chl-*a*, and that the only watershed characteristics that revealed a significant positive correlation were watershed area and the ratio of watershed area to waterbody depth. In the global study by Brylinsky and Mann (1973), however, only ~2% variation of phytoplankton production was explained by watershed area in regression models. Hence, our literature review indicates that no single watershed characteristic consistently serves as the best predictor of Chl-*a* within waterbodies at either global or regional scales.

With the development of remote sensing and geographic information systems (GIS), evaluating which watershed characteristics influence Chl-*a* in waterbodies at a broad spatial scale is now possible. New technology using satellite imagery allows scientists to estimate Chl-*a* at global scales (e.g., Sayers et al. 2015). Additionally, global datasets for variables that potentially influence Chl-*a* in waterbodies are now available for land cover (Arino et al. 2012), phosphorus application (Potter et al. 2011), and climate (Hijmans et al. 2005), and most HGM variables such as watershed area and total stream length can now feasibly be calculated using GIS (Turcotte et al. 2001). More important, these technological advances allow us to gather information in remote and understudied regions where *in situ* data are lacking; however, utilizing remote sensed data to characterize watershed characteristics is still limited by the need to delineate watershed and shoreline boundaries in GIS data layers.

In this study, we delineated watersheds and shorelines for 227 waterbodies distributed across 5 continents and compiled available data to evaluate the hypothesis that watershed characteristics could be used to predict Chl-*a* in downstream waterbodies at the global scale. We built a conceptual model (elaborated later) in which we considered how characteristics of the watershed and waterbody may influence Chl-*a*. Note that our model was limited to select variables that we could acquire for all waterbodies at the global scale. We then developed a statistical model to determine which of our remotely sensed factors most parsimoniously explained variation in Chl-*a* in waterbodies. To explore whether the variation of Chl-*a* can be better explained at the regional scale, we grouped the waterbodies based on the Freshwater Ecoregions of the World (FEOW; Abell et al. 2008) and applied the same statistical modeling approach to the FEOW with the largest sample size, the Laurentian Great Lakes region.

**Table 1**. Description of variables used in this study and information for data sources.

| No. | Variable | Unit | Source |
|---|---|---|---|
| | Chlorophyll *a* in waterbody ($Chl_{lake}$) | µg L$^{-1}$ | Sayers et al. (2015) |
| | **Nutrient condition** | | |
| 1 | Urban land cover (%$_{urb}$) | % | Arino et al. (2012) |
| 2 | Agricultural land cover (%$_{ag}$) | % | Arino et al. (2012) |
| 3 | Undeveloped land cover (%$_{undev}$) | % | Arino et al. (2012) |
| 4 | Snow land cover (%$_{snow}$) | % | Arino et al. (2012) |
| 5 | Water land cover (%$_{water}$) | % | Arino et al. (2012) |
| 6 | Total phosphorus applied ($TP_a$) | kg km$^{-2}$ | Potter et al. (2011) |
| 7 | Human population (Pop) | human km$^{-2}$ | CIESIN (2004) |
| | **Hydrogeomorphic (HGM) characteristics** | | |
| 8 | Waterbody area ($A_{lake}$) | km$^2$ | This study |
| 9 | Shoreline development index (SDI) | | This study |
| 10 | Hydrological residence time index (HRT) | cells km$^{-2}$ | This study |
| 11 | Watershed area ($A_{shed}$) | km$^2$ | This study |
| 12 | Watershed area to waterbody area ratio ($A_{shed}$:$A_{lake}$) | | This study |
| 13 | Watershed compactness coefficient (WCC) | | This study |
| 14 | Total stream length ($L_{str}$) | km | This study |
| | **Regional climate** | | |
| 15 | Annual precipitation ($P_{mean}$) | mm | Hijmans et al. (2005) |
| 16 | Precipitation of wettest quarter ($P_{wet}$) | mm | Hijmans et al. (2005) |
| 17 | Precipitation of driest quarter ($P_{dry}$) | mm | Hijmans et al. (2005) |
| 18 | Precipitation seasonality (coefficient of variation, $P_{season}$) | % | Hijmans et al. (2005) |
| 19 | Mean annual temperature ($T_{mean}$) | °C | Hijmans et al. (2005) |
| 20 | Mean temperature of the warmest quarter ($T_{hot}$) | °C | Hijmans et al. (2005) |
| 21 | Mean temperature of the coldest quarter ($T_{cold}$) | °C | Hijmans et al. (2005) |
| 22 | Temperature seasonality (standard deviation, $T_{season}$) | °C | Hijmans et al. (2005) |
| 23 | Annual mean insolation (Insol) | kWh m$^{-2}$ d$^{-1}$ | NASA (2008) |
| 24 | Precipitation of the warmest quarter ($P_{hot}$) | mm | Hijmans et al. (2005) |
| 25 | Precipitation of the coldest quarter ($P_{cold}$) | mm | Hijmans et al. (2005) |
| 26 | Mean temperature of the wettest quarter ($T_{wet}$) | °C | Hijmans et al. (2005) |
| 27 | Mean temperature of the driest quarter ($T_{dry}$) | °C | Hijmans et al. (2005) |

## Methods

### Conceptual model

We identified 27 variables (Table 1) of watershed and waterbody characteristics that can be estimated remotely and influence the availability of nutrients and, consequently, Chl-*a*. We categorized variables into 5 types: (a) nutrient condition of the watershed, (b) HGM characteristics of the watershed, (c) HGM characteristics of the waterbody, (d) regional climate influences in the watershed, and (e) regional climate influences in the waterbody. Then we built a conceptual model to describe how these variables may influence the Chl-*a* of a waterbody (Fig. 1).

The nutrient condition of a watershed strongly influences the nutrient availability to the receiving waterbody. We collected data on types of land cover (Table 1, No. 1–5) and phosphorus application (Table 1, No. 6) within the watershed, both of which have been linked to the trophic state of the waterbody (Allen and Kramer 1972, Carpenter et al. 1998, Knoll et al. 2003,

Bremigan et al. 2008). Note that "snow" and "water" among land cover types (Table 1) represented percentages of watershed area covered by snow (and ice) and waterbodies that generally generate no nutrient input to the receiving waterbody, respectively (Arino et al. 2012). We also included human population (Table 1, No. 7) in this analysis, which is positively related to nutrient availability to the waterbody (Weibel 1969, Cole et al. 1993).

HGM characteristics of a watershed can influence Chl-*a* in a downstream waterbody by determining the quantity and delivery rate of nutrients. These characteristics include watershed area, the ratio of watershed to waterbody area, the shape of the watershed, and total stream length within the watershed (Table 1, No. 11–14). Watershed area and the ratio of watershed to waterbody area are both positively related to the quantity of nutrients available to the waterbody (Black 1972, Knoll et al. 2003, Bremigan et al. 2008). The shape of the watershed in this study was indicated by the watershed compactness coefficient (WCC; Black 1972). A lower WCC indicates a more circular watershed with shorter traveling distances for nutrients to reach the waterbody. Total stream length in a watershed is an indicator of hydrologic connectivity, which is positively related to nutrient inputs to the waterbody (Gordon et al. 1992).

Once nutrients reach a waterbody, its own HGM characteristics also influence Chl-*a*. Our analysis included area (Table 1, No. 8); shape, measured as the shoreline development index (SDI, Table 1, No. 9; Wetzel 2001); and an index of hydrological residence time (Table 1, No. 10). Waterbodies with longer hydrological residence time

may have lower Chl-*a* because of lower nutrient replenishing rates (Wetzel 2001). Area is generally positively correlated to volume and thus the hydrological residence time (defined later; Nõges 2009). Waterbodies with higher SDI generally have a greater percentage of littoral (higher productivity) than open water (lower productivity) area. Thus, waterbodies with larger areas and lower SDI values may have lower Chl-*a* (Wetzel 2001).

Aquatic systems are also subject to climatic controls on Chl-*a* that occur in the watershed. We included several measures of precipitation (Table 1, No. 15–18) that can influence Chl-*a* through different ways. Increases in precipitation may lead to increases in nutrient inputs and episodic increases in Chl-*a* in a waterbody (Soranno et al. 1996, 1997, Chambers et al. 2006). By contrast, increases in precipitation may also lead to higher water levels, resulting in a dilution of the nutrient concentration and subsequent reductions in Chl-*a* (Markensten 2006, Sass et al. 2008).

Regional climate conditions also influence Chl-*a* within the waterbody. Climate influences on Chl-*a* can be measured via temperature (Table 1, No. 19–22) and insolation (Table 1, No. 23), both of which influence phytoplankton growth directly and indirectly (Wetzel 2001). Variables indicating interactions between precipitation and temperature may also be important (Table 1, 24–27). For example, changes in spring temperatures, when rain events are most common in temperate regions, can have a disproportionally large influence on phytoplankton growth within the growing season (Weyhenmeyer 2001, Sass et al. 2008).
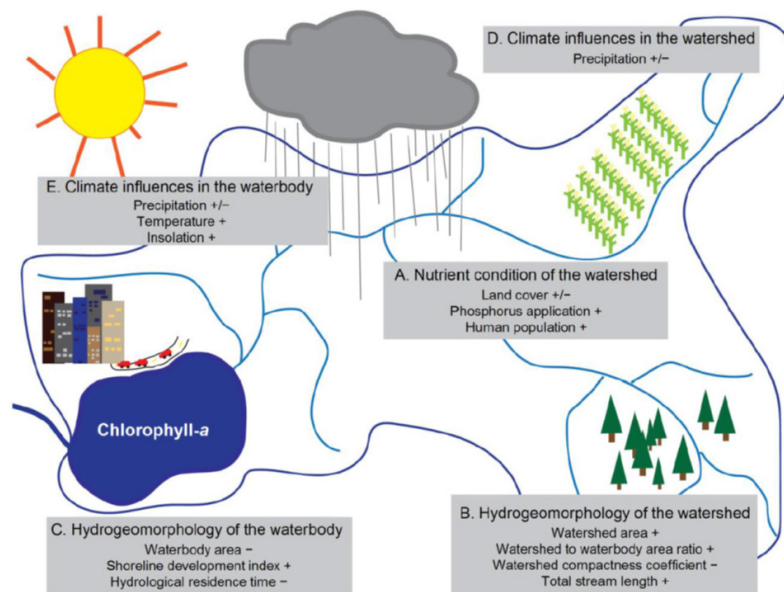


**Fig. 1**. Conceptual model describing how watershed (a and b), waterbody (c), and climate (d and e) characteristics may positively (+) or negatively (−) influence chlorophyll *a* in the waterbody.

## Delineation of global watersheds and waterbodies

We used the procedure outlined in Merwade (2012) and Maidment (2002) to delineate watersheds of 227 waterbodies (Fig. 2). Watersheds were delineated based on National Aeronautics and Space Administration's (NASA) Advanced Spaceborne Thermal Emission and Reflection Radiometer Global digital elevation model version 2 (ASTER GDEM 2, available at http://asterweb. jpl.nasa.gov/gdem.asp) and the HydroSHEDS stream network layer (Lehner et al. 2008). For large watersheds of Lake Winnipeg and Lake Nasser that were beyond our computational capacity, we obtained data from Agriculture and Agri-Food Canada and the Transboundary Waters Assessment Programme, respectively. Geospatial shoreline data on these 227 waterbodies were obtained from the Global Lakes and Wetlands Database Level 2 (GLWD-2; Lehner and Doll 2004) or manually delineated from high-resolution imagery in a GIS software ArcGIS (http://www.esri.com/software/arcgis/).

Our global dataset included watersheds and waterbodies across North America ($n = 180$; $n = 51$ within the Laurentian Great lakes region, according to FEOW), South America ($n = 8$), Asia ($n = 3$), Australia ($n = 2$), Europe ($n = 24$), and Africa ($n = 10$), ranging from ~65°N to 35°S. Waterbodies in the dataset ranged from $4 \times 10^{-3}$ km$^2$ to $2.1 \times 10^{11}$ km$^2$ in area, from 27.7 km$^2$ to $9.8 \times 10^{14}$ km$^2$ in watershed area, and from 0.9 μg L$^{-1}$ to 66.9 μg L$^{-1}$ in growing season Chl-*a*.

## Chl-*a* in waterbodies

Chl-*a* data included both remotely sensed estimates ($n = 221$) and *in situ* measurements ($n = 6$). Remotely sensed Chl-*a* estimates were taken from Sayers et al. (2015), representing averages in the 2011 growing season (August in Northern Hemisphere, February in Southern Hemisphere). For waterbodies without remotely sensed estimates, we used available *in situ* Chl-*a* measurements in the growing season of a year closest to 2011 from literature or the US Environmental Protection Agency (EPA) STORET database (http://www.epa.gov/storet/).

## Nutrient condition of the watershed

To derive variables of nutrient conditions (Table 1, No. 1–7), we used the package "rgdal" (Keitt et al. 2010) in R (R Core Team 2015) or ArcGIS to extract data for the extent of each watershed from GIS data layers. For land cover variables, we classified 23 land cover types in the European Space Agency's GLOBCover dataset (Arino et al. 2012) into 5 general categories and calculated the proportion of watershed area covered by each category (Table 1, No. 1–5). We used the most recent 2000 population data (CIESEN 2005) from NASA's Socioeconomic Data and Applications Center (SEDAC) to calculate human population density (Pop). Although these data were collected considerably earlier than our Chl-*a* data, this dataset is the most recent and reliable to our knowledge. Phosphorus application data were from
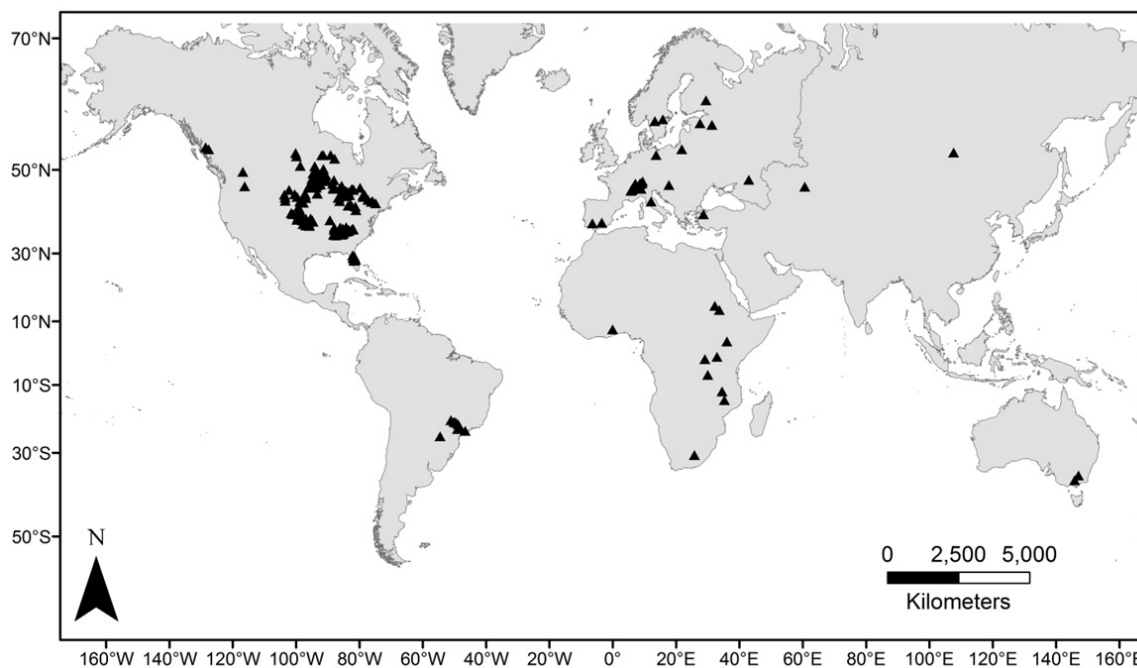


**Fig. 2**. Geographic distribution of 227 waterbodies in this study.

estimated phosphorus contents of recorded manure and fertilizer applications between 1994 and 2000 (Potter et al. 2010, 2011). We calculated the total phosphorus applied ($TP_a$) as the total amount of phosphorus in manure production and fertilizer application per unit area.

## HGM characteristics

We calculated several HGM variables (Table 1, No. 8–14) based on the shoreline and watershed GIS data layer developed in this study. We used the XTools Package 11.1 in ArcGIS (Data East, LLC) to calculate waterbody area ($A_{lake}$) and watershed area ($A_{shed}$). We calculated SDI and the WCC as (Wetzel 2001):

$$SDI = 0.5 \times S_{lake} / (\pi \times A_{lake})^{0.5} \text{ and} \qquad (1)$$

$$WCC = 0.5 \times S_{shed} / (\pi \times A_{shed})^{0.5}, \qquad (2)$$

where $S_{lake}$ and $S_{shed}$ are the perimeters of the waterbody and the watershed, respectively. We used a hydrological residence time index (HRT) as a proxy for hydrological residence time, defined as flow accumulation cells (a proxy of discharge) divided by $A_{lake}$ (a proxy of waterbody volume). Flow accumulation cells were calculated as the total number of ASTER GDEM 2 digital elevation model raster cells in the watershed that flow into the waterbody pour point. We extracted total stream length in each watershed ($L_{Str}$) from HydroSHEDS stream network layer.

## Regional climate

Regional climate variables (Table 1, No. 15–27) were derived using the WorldClim Global Climate dataset for 1950–2000 (Hijmans et al. 2005) and data from NASA's Surface Meteorological and Solar Energy web portal (https://eosweb.larc.nasa.gov/sse/). We extracted data for the extent of a watershed using the package "rgdal" in R or ArcGIS.

## Statistical analysis

We used a regression approach to relate Chl-*a* in global waterbodies ($Chl_{lake}$) to variables representing watershed, waterbody, and regional climate characteristics (Table 1). Our initial analyses involved examining the scatterplot between $Chl_{lake}$, the response variable, and each potential predictor variable. We found that relationships between $Chl_{lake}$ and most predictor variables were nonlinear, suggesting that the generalized additive model (GAM; Hastie and Tibshirani 1990) is an appropriate tool because predictor variables can be incorporated as nonlinear smooth functions in the model. We also found that the dis-

tributions of 6 variables ($A_{lake}$, $A_{shed}$, $A_{shed}$:$A_{lake}$, Pop, HRT, and $L_{str}$) were highly skewed, so they were log-transformed before the GAM analysis.

The general form of a GAM can be expressed as:

$$g[E(Chl_{lake})] = \beta_0 + \Sigma_i \, s_i(x_i, \, df_i), \qquad (3)$$

where $g[E(Chl_{lake})]$ is a link function that relates the expected value of $Chl_{lake}$ to predictor variables $x_i$; $\beta_0$ is the intercept; and $s_i(x_i, \, df_i)$ is the "additive predictor," a nonlinear smoothing function of the predictor variable $x_i$ with $df_i$ degrees of freedom. The selection of link function depends on the probability distribution used in GAMs. In our initial analyses, we found that the spread of $Chl_{lake}$ increased disproportionally with changes in most of potential predictor variables, suggesting a gamma distribution was appropriate. The 3 commonly used link functions for gamma GAMs are reciprocal, log, and identity links. We used the log-link function to ensure that modeled $Chl_{lake}$ values were always positive.

We used a forward model selection approach to identify the variables that best explained the variation of $Chl_{lake}$ among global waterbodies ($n = 227$) and the package "mgcv" (Wood 2015) in R to perform the GAM analysis. We estimated smoothing functions nonparametrically using penalized regression spline methods, in which the degrees of freedom of each smoothing function are objectively estimated based on the fit to empirical data. We used cubic regression splines and the restricted maximum likelihood (REML) estimation for fitting GAMs (Wood 2015), a robust method for variable selection (Marra and Wood 2011).

To find the most parsimonious model of global waterbodies, we used Akaike information criterion (Akaike 1974) with a correction for finite sample sizes ($AIC_c$; Burnham and Anderson 2002). In each step to add an additive predictor, we calculated $\Delta AIC_c$ for every potential model as the difference between its $AIC_c$ and the minimum $AIC_c$ among models up to this step. We filtered out models with $\Delta AIC_c > 2$ because they were unlikely to fit the data better than the model with minimum $AIC_c$ (Burnham and Anderson 2002). After the selection of a predictor variable, we removed highly correlated predictor variables ($|r| > 0.7$) from the potential variable pool before entering next step. We repeated this forward-selection step until adding an additive predictor to a selected model failed to reduce $AIC_c$ by >2.

We used the same model selection method to find the most parsimonious Great Lakes regional model ($n = 51$). To further explore the benefit of developing regional models, we also compared the most parsimonious global model with a Great Lakes regional model created with the same selected variables as in the global model.

**Table 2**. Selected results for forward generalized additive model selection for the global model of chlorophyll *a* in waterbodies (Chl$_{lake}$). Refer to Table 1 for variable abbreviations.

| Model | Deviance explained | $\Delta AIC_c$ | Note |
|---|---|---|---|
| Chl$_{lake}$ ~ A$_{lake}$ + T$_{hot}$ + SDI + A$_{shed}$:A$_{lake}$ | 30.9% | 0.0 | Best model |
| Chl$_{lake}$ ~ A$_{lake}$ + T$_{hot}$ + SDI | 28.4% | 1.8 | Best 3-variable model |
| Chl$_{lake}$ ~ A$_{lake}$ + T$_{hot}$ + HRT | 27.2% | 3.1 | Log-transformed HRT is highly correlated to log-transformed A$_{shed}$:A$_{lake}$ ($r = 0.97$). |
| Chl$_{lake}$ ~ A$_{lake}$ + T$_{hot}$ + A$_{shed}$:A$_{lake}$ | 27.0% | 3.4 | |
| Chl$_{lake}$ ~ A$_{lake}$ + T$_{hot}$ | 24.4% | 5.4 | Best 2-variable model |
| Chl$_{lake}$ ~ A$_{lake}$ | 17.0% | 17.9 | Best 1-variable model |

## Results

### Global model

Our final selected GAM for global waterbodies explained 30.9% of the deviance (Table 2) and included 4 additive predictors: A$_{lake}$, mean temperature of the warmest quarter (T$_{hot}$), SDI, and A$_{shed}$:A$_{lake}$. Another model that could be equally predictive ($\Delta AIC_c$ = 1.8) included 3 of 4 additive predictors in the final selected model (A$_{lake}$, T$_{hot}$, and SDI) and explained 28.4% of deviance (Table 2). Relationships between Chl$_{lake}$ and each of A$_{lake}$, T$_{hot}$, and SDI in these 2 GAMs were similar.

In the final selected GAM for global waterbodies, Chl$_{lake}$ was nonlinearly related to all 4 predictor variables (Fig. 3). Chl$_{lake}$ decreased with increases in A$_{lake}$, as predicted in the conceptual model (Fig. 2), when A$_{lake}$ was

less than ~103 km$^2$; however, Chl$_{lake}$ increased with A$_{lake}$ when A$_{lake}$ was less than ~10 km$^2$ and was almost unchanged with A$_{lake}$ when A$_{lake}$ was between ~10 and $10^3$ km$^2$. Chl$_{lake}$ increased with T$_{hot}$, as predicted in the conceptual model, but asymptotically. Further increases in T$_{hot}$ above 22 °C led to almost no change in Chl$_{lake}$. The relationship between Chl$_{lake}$ and SDI was polynomial-like, with the coefficient varying nonlinearly around zero over most of the SDI gradient. The most unambiguous aspect of the relationship occurred when SDI was less than ~3, where Chl$_{lake}$ decreased with SDI, opposite to the prediction in the conceptual model. The effect of A$_{shed}$:A$_{lake}$ seemed to be driven by waterbodies at the extreme ends of the gradient. When A$_{shed}$:A$_{lake}$ was relatively small (less than ~30), an increase in the ratio tended to produce higher estimates of Chl$_{lake}$, as predicted in the conceptual model; however, the response in Chl$_{lake}$ reversed as the A$_{shed}$:A$_{lake}$ ratio continued to increase.
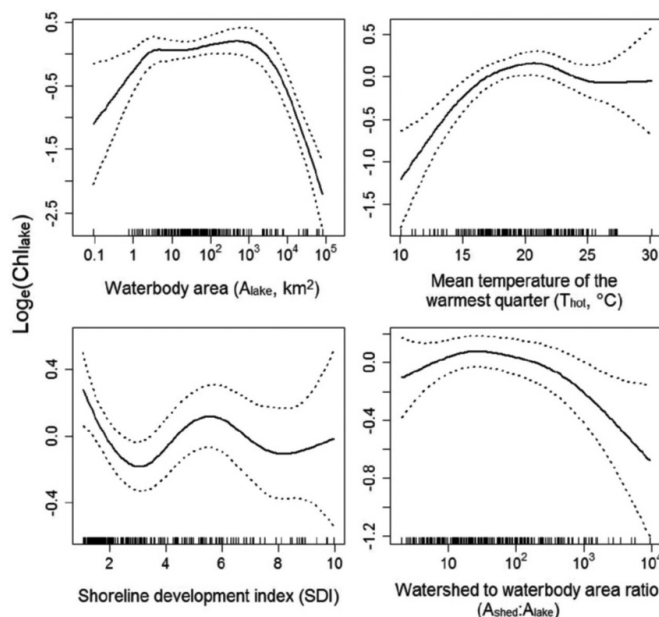


**Fig. 3**. The relationship between each predictor variable and chlorophyll *a* in 227 global waterbodies in the generalized additive model selected based on Akaike information criterion (Table 2). The solid line represents the estimated nonlinear smooth function of a predictor variable. Dotted lines represent the ~95% confidence interval. Dashes above the horizontal axis represent the distribution of data points.
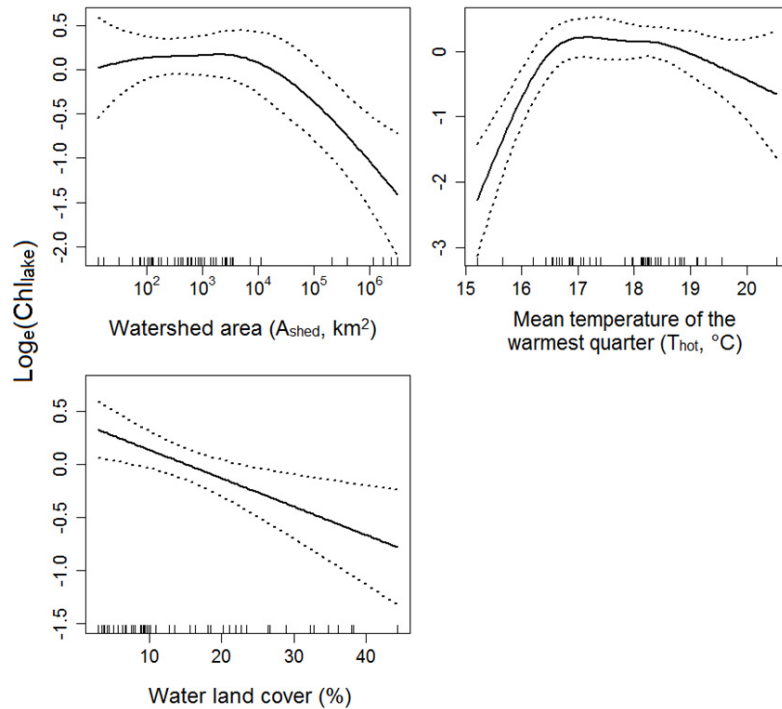
**Fig. 4**. The relationship between each predictor variable and chlorophyll *a* in 51 waterbodies of the Laurentian Great Lakes region in the generalized additive model selected based on Akaike information criterion (Table 3). The solid line represents the estimated nonlinear smooth function of a predictor variable. Dotted lines represent the ~95% confidence interval. Dashes above the horizontal axis represent the distribution of data points.
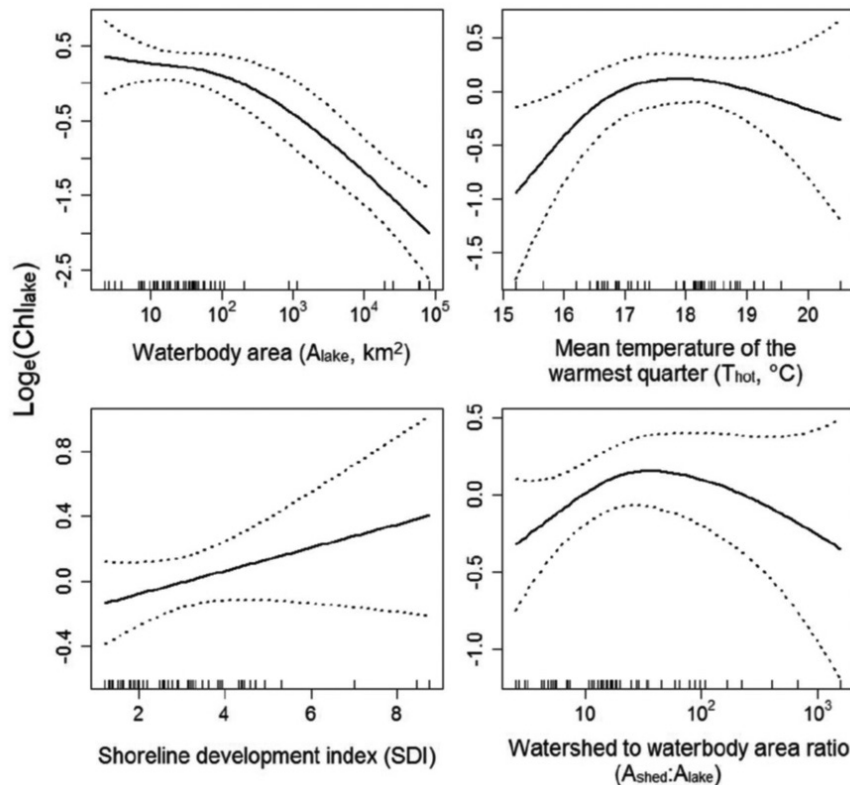


**Fig. 5**. The relationship between each predictor variable selected in the global model (Fig. 3) and chlorophyll *a* in 51 waterbodies of the Laurentian Great Lakes region based on the generalized additive model. The solid line represents the nonlinear smooth function of a predictor variable. Dotted lines represent the ~95% confidence interval. Dashes above the horizontal axis represent the distribution of data points.

**Table 3**. Selected results for forward generalized additive model selection for the Laurentian Great Lakes regional model of Chlorophyll *a* in waterbodies ($Chl_{lake}$). Refer to Table 1 for variable abbreviations.

| Model | Deviance explained | $\Delta AIC_c$ | Note |
|---|---|---|---|
| $Chl_{lake} \sim A_{shed} + T_{hot} + \%_{water}$ | 58.4% | 0.0 | Best model |
| $Chl_{lake} \sim A_{shed} + T_{hot}$ | 56.1% | 2.7 | Best 2-variable model |
| $Chl_{lake} \sim A_{lake}$ | 37.4% | 7.4 | Best 1-variable model $A_{lake}$, $A_{shed}$, and $L_{str}$ are highly correlated after log-transformation ($r > 0.87$). |
| $Chl_{lake} \sim A_{shed}$ | 37.8% | 7.7 | |
| $Chl_{lake} \sim L_{str}$ | 37.6% | 7.8 | |
| $Chl_{lake} \sim A_{lake} + T_{hot} + SDI + A_{shed}{:}A_{lake}$ | 49.5% | 12.9 | Best global model |

## Laurentian Great Lakes regional model

Our final selected GAM for waterbodies in the Great Lakes region explained 58.4% of deviance (Table 3) and included $A_{shed}$, $T_{hot}$, and water land cover (%water). $T_{hot}$ was the only common variable selected between the global and regional models; however, $A_{shed}$ (selected in regional model) and $A_{lake}$ (selected in global model) were highly correlated after log-transformation within either global ($r = 0.78$) or regional ($r = 0.87$) waterbodies.

In the final selected GAM for waterbodies in the Great Lakes region, $Chl_{lake}$ was nonlinearly related to $A_{shed}$ and $T_{hot}$ but decreased linearly with $\%_{water}$ (Fig. 4). $Chl_{lake}$ increased slightly with $A_{shed}$ when $A_{shed}$ was less than ~$10^4$ km², as predicted by the conceptual model (Fig. 2), but decreased sharply with $A_{shed}$ when $A_{shed}$ was less than ~$10^4$, which was influenced by extremely large watersheds (i.e., the Laurentian Great Lakes). Similar to the global model, $Chl_{lake}$ increased asymptotically with $T_{hot}$ but only up to ~17 °C (compared to ~22 °C in the global model). At warmer temperatures, however, $Chl_{lake}$ decreased with increasing temperatures but with wider confidence intervals.

The Great Lakes regional GAM created with the same selected variables as in the final selected global model explained 49.5% of the deviance but had an $\Delta AIC_c$ of 12.6 (Table 3). The shape of the relationships between the selected variables and $Chl_{lake}$ in the global (Fig. 3) and regional (Fig. 5) models were largely similar, with the exception of SDI. In the regional model, $Chl_{lake}$ increased linearly with SDI, consistent with the prediction in the conceptual model.

## Discussion

Our results showed that ~30% of Chl-*a* variation across global waterbodies could be explained by 3 HGM characteristics ($A_{lake}$, SDI, and $A_{shed}{:}A_{lake}$) and a climate variable ($T_{hot}$). Predicting phytoplankton production and Chl-*a* at

the global scale is uncommon; we are aware of only 2 other studies. First, Brylinsky and Mann (1973) found that mean annual air temperature had the strongest effects on phytoplankton production across 55 global waterbodies, whereas nutrient and morphological variables were less important. Our model also showed the importance of temperature to Chl-*a*. In addition, the variable mean $T_{hot}$ in our global model was highly correlated with mean annual temperature ($r = 0.82$). Mean annual air temperature alone, however, could explain 59% of variation in Brylinsky and Mann's (1973) model, which is much higher than variation explained by our model. Second, Meeuwig and Peters (1996) studied 63 global lakes and developed an empirical land-use model that included residence time, human population in the watershed, forested land cover, and mean waterbody depth and were able to explain 55–72% of the variance in Chl-*a*. In a relatively large-scale regional study of 1337 European lakes, Nõges (2009) reported the ratio of watershed area to waterbody depth to have the strongest correlation with Chl-*a* ($r \sim 0.3$), similar to the variance explained by our model. Together, these studies suggest that improved waterbody-specific information, such as depth and hydrological residence time, should be included in future Chl-*a* modeling efforts at global scales.

Our global model revealed a largely negative relationship between $A_{lake}$ and Chl-*a*. We were unable to obtain depth and volume information for all waterbodies, which have been commonly used to explain variation of Chl-*a* in lakes since Vollenweider's (1975) phosphorus-loading–Chl-*a* model. Both depth and volume, however, generally increase with $A_{lake}$ (Håkanson 2005). Thus our results are consistent with previous studies that found lake trophic status to decline with increasing depth (Meeuwig and Peters 1996, Bremigan et al. 2008, Liu et al. 2011). Notably, the dataset from Meeuwig and Peters (1996) was dominated by large lakes (mean $A_{lake} = 1.8 \times 10^3$ km²), which is approximately where our model begins to show a downward trend in the relationship between Chl-*a* and

$A_{lake}$ (Fig. 3). Overall, these findings indicate that waterbody area may serve as a proxy for depth or volume when this information is not available, but that the inclusion of depth or volume data may improve the model.

In both the global and the Great Lakes regional models, $T_{hot}$ was selected as an important predictor and showed a similar relationship with Chl-*a* regionally and globally. $T_{hot}$ describes basic growing conditions for Chl-*a* and is commonly cited as a strong control on phytoplankton in waterbodies (Weyhenmeyer 2001, Markensten 2006, Blenckner et al. 2007). Further, our model indicated a threshold beyond which increases in $T_{hot}$ do not result in clear increases in Chl-*a*, although such thresholds can differ by region. The asymptotic relationship between $T_{hot}$ and Chl-*a* suggested that temperature was not likely limiting Chl-*a* production at higher $T_{hot}$ levels, but production could be limited by other factors such as nutrient availability or euphotic depth (e.g., Jones et al. 2011). This result suggests that Chl-*a* is more sensitive to changes in $T_{hot}$ when $T_{hot}$ is relatively cool. Further, because $T_{hot}$ was negatively correlated with latitude ($r = -0.68$ in our global dataset), our results also imply that temperate waterbodies may be more sensitive to a warming climate than tropical waterbodies.

Although the SDI was included in the final models at the global scale, the relationship with Chl-*a* is not easy to interpret and could be the result of nonlinear overfitting. The region with the tightest confidence intervals suggests that Chl-*a* declines with increasing SDI up to SDI values of ~3 (Fig. 3), values that include waterbodies with relatively little shoreline length for a given area or a more circular shape. This pattern is opposite from our prediction that waterbodies with more littoral habitat would be more productive. This positive (and linear) relationship between SDI and Chl-*a* was revealed when SDI was included as a predictor variable with only data from the Laurentian Great Lakes region. These conflicting relationships suggest the pattern derived in the global model is spurious, and future work will be required to determine whether SDI can be a useful predictor for Chl-*a* at broad spatial scales.

The last selected variable in our global model, $A_{shed}{:}A_{lake}$, unexpectedly exhibited a negative nonlinear relationship with Chl-*a*. Our conceptual model proposed that an increasing ratio would lead to higher Chl-*a*, developed from previous studies (Håkanson 2005, Bremigan et al. 2008), albeit with waterbodies that were generally smaller than those included in our dataset. One potential explanation is that as $A_{shed}{:}A_{lake}$ increases, connected upstream waterbodies and wetlands are more likely to be present within the watershed and may intercept nutrients before they reach downstream (see Soranno et al. 1999, 2015). Additionally, an increase in

wetland area within a watershed may also result in declining Chl-*a*, as seen in previous work showing an inverse relationship between the nutrient status of a waterbody and wetland area within the watershed (Detenbeck et al. 1993, Weller et al. 1996, Reed and Carpenter 2002). Greater wetland area could also limit production by contributing colored compounds to the waterbody (Martin et al. 2011), decreasing water clarity and causing light limitation for phytoplankton.

Our Laurentian Great Lakes regional model explained substantially more Chl-*a* variation across waterbodies (58%) than the global model (31%). In contrast to our global model, the regional model included $A_{shed}$ as an HGM variable. The modeled relationship between Chl-*a* and $A_{shed}$ was weakly positive for a large number of small watersheds, consistent with the prediction from our conceptual model. One possible explanation for this weak relationship is that positive influences of increasing $A_{shed}$ on Chl-*a* might be offset by negative influences of increasing $A_{lake}$ in our dataset because they were highly positively correlated after log-transformation ($r = 0.87$). The modeled relationship between Chl-*a* and $A_{shed}$ was strongly negative for a small number of large watersheds, including watersheds of the 5 Great Lakes and Lake St. Clair, opposite from the prediction of the conceptual model. Because we calculated the watershed of each waterbody as containing watersheds of all upstream waterbodies, the calculated $A_{shed}$ for some large waterbodies may not be indicative of their true nutrient availability. A large proportion of nutrients flushed into these would not be transported to the other waterbodies because they may have retention times >50 years (USEPA and EC 1995). For example, Lake Ontario, the most downstream Great Lake, has the largest watershed because it contains the entire Great Lakes basin in our analysis; however, the bulk of available nutrients from the Lake Ontario watershed are likely utilized in upstream waterbodies before reaching the lake.

The last selected variable in the Great Lakes regional model, the percent of water area in the watershed (%$_{water}$), was negatively related to Chl-*a*, consistent with the prediction from our conceptual model. Because watershed area covered by waterbodies generally would not generate nutrient input to the receiving waterbody, a higher value of %$_{water}$ may indicate a lower nutrient condition of the watershed. Adding this variable to the model with $A_{lake}$ and $T_{hot}$, however, could explain only ~2% more variation in Chl-*a* across waterbodies in the Great Lakes region (Table 3), indicating that traditional nutrient indicators of the watershed (land cover, phosphorus application, and human population) were not informative to predicting Chl-*a* in either our global or regional datasets. Brylinsky and Mann (1973) also found that nutrient concentration

variables (e.g., TP and total nitrogen) were relatively less important to phytoplankton production in waterbodies at a global scale, but the importance of nutrient concentration variables increased for a subset of lakes between 39°N and 55°N.

We believe our results differ from many studies that highlighted the importance of land cover because of issues associated with the limited land cover gradients in our dataset. For example, many of the watersheds had a low percentage of agricultural land cover (mean $\%_{ag}$ = 8.5%), possibly due to the geographical distribution of our waterbodies. Even so, we are not the first study to find no linkage between Chl-*a* and $\%_{Ag}$ (see Knoll et al. 2003, Burford et al. 2007, Carneiro et al. 2014). Knoll et al. (2003) noted that land cover is not a reliable indicator of nutrient loading because of variation in actual nutrient export rates. Although we attempted to include other proxies of nutrient loading (TP$_a$ and Pop), they also failed to explain variation in Chl-*a*. Our model may benefit from including more waterbodies that cover a larger gradient of percent agricultural land cover in watersheds, or even improved estimates of land cover or phosphorus application not currently feasible at the global scale. Similarly, future models that incorporate nitrogen inputs also could be helpful given that nitrogen can also be limiting for primary production in inland lakes (Elser et al. 2007). Incorporating nutrients in broad-scale Chl-*a* models should consider the spatial extent of watersheds and land use within them, lake connectivity, and region effects (Sorrano et al. 2015). Abell et al. (2012) warned that the ability to predict Chl-*a* based on nutrient measures is less reliable in tropical and polar lakes, which further suggests the importance of considering region for large-scale evaluations.

One additional possible reason for the relatively low explanatory power of our global model was our inability to account for biological processes occurring within waterbodies. For example, our model did not account for food-web constraints on phytoplankton. Inclusion of herbivory rates may offer improved insight into Chl-*a* dynamics, as has been shown in various studies (e.g., Vanni et al. 2005, Gorman et al. 2014). Further, food-web constraints on phytoplankton might be more consistent at a regional scale but not at the global scale. For example, in the Laurentian Great Lakes region, invasive dreissenid mussels are known to be extremely effective at filtering Chl-*a* (Guidford et al. 2013) and sequestering nutrients, thereby changing the nutrient profile (Hecky et al. 2004).

Another limitation of our study was our reliance on data that were not temporally coincident with our Chl-a dataset. First, our climate indices derive from long-term averages, whereas our Chl-*a* data are month-long snapshots from a single year, leaving room for a mismatch between climate signals and Chl-*a* estimates in our dataset. Second, in eutrophic waterbodies (Chl$_{lake}$ > 14.3 µg L$^{-1}$; Wetzel 2001), which describes many of ours (*n* = 159), the timing of weather events can be more important than long-term climate metrics to certain kinds of epilimnetic phytoplankton (Soranno et al. 1997). Third, whereas our Chl-*a* data were from 2011, our population data represented conditions in 2000 and our phosphorus data represented conditions in 1994–2001. These data are possibly no longer representative of the conditions in our waterbodies when the Chl-*a* data were extracted. Depending on changes in agricultural land cover, this may or may not accurately characterize contemporary conditions.

Issues with data integrity or consistency can be mitigated by pursuing a regional rather than a global approach, which our results have shown explain more variance when predicting Chl-*a*. Data for variables not included in our analysis are also more likely to be available at the regional scale and may add important information to predicting Chl-*a*. Although we included proxies for variables such as waterbody depth, volume, discharge, and hydrological residence time, our model would likely be improved if direct empirical data could be obtained for these variables because they are crucial to basic limnological models estimating nutrient concentrations (Dillon and Rigler 1974, Vollenweider 1975). Regional geology has also been shown to be important to Chl-*a* in waterbodies because it controls the ambient nutrients available for transport (Dillon and Kirchner 1975, Jones and Bachmann 1977, Canfield and Hoyer 1988). Future models could consider underlying geology within the watershed, in addition to land cover, as another predictor variable. Finally, although a majority of studies categorize land cover variables into the groups we used, more nuanced land-cover categories might be more helpful. Parsing out details such as the proportion of wetland cover in the watershed (Devito et al. 2000, Martin et al. 2011), the type of agricultural land use (Knoll et al. 2003), and metrics assessing the cohesion of land cover categories (Gemesi et al. 2011) have all been shown to be important to different water quality measures.

Overall, our results indicate that a regional approach to watershed management can best predict Chl-*a* in waterbodies. In regions of the world where a regional approach is not feasible, our global model could still rely on remotely sensed information to make predictions about Chl-*a*, but the confidence intervals around that prediction would be relatively wide. Based on our synthesis of many regional studies, there are also regional differences in the variables that best explain variation in Chl-*a*. Our results failed to identify major watershed influences, and most of

the key predictor variables (surface area, temperature, and watershed size), cannot be manipulated by managers. The variables that managers can most easily influence are agricultural and urban land covers and nutrient application within the watershed, and neither of these were important predictors for our global and regional models. The lack of direct watershed influences may be due to a lack of high-quality nutrient indicator data or the limited gradient for nutrient inputs that occurred in our dataset. In conclusion, our results provided a global model to coarsely predict Chl-*a* in waterbodies where only remotely sensed data can be easily acquired and suggest that more accurate models can be developed at regional scales that would better inform management and policy decisions.

## Acknowledgements

## References

Abell JM, Özkundakci D, Hamilton DP, Jones JR. 2012. Latitudinal variation in nutrient stoichiometry and chlorophyll–nutrient relationships in lakes: a global study. Fund Appl Limnol. 181:1–14.

Abell R, Thieme ML, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Balderas SC, Bussing W, et al. 2008. Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. Bioscience. 58:403–414.

Akaike H. 1974. A new look at the statistical model identification. IEEE T Automat Contr. 19:716–723.

Allen HE, Kramer JR. 1972. Nutrients in natural waters: environmental science and technology series. New York (NY): Wiley.

Arino O, Ramos Perez JJ, Kalogirou V, Bontemps S, Defourny P, Van Bogaert E. 2012. Global land cover map for 2009 (GlobCover 2009). European Space Agency and Université Catholique de Louvain (Belgium).

Black PE. 1972. Hydrograph responses to geomorphic model watershed characteristics and precipitation variables. J Hydrol. 17:309–329.

Blenckner T, Adrian R, Livingstone DM, Jennings E, Weyhenmeyer GA, George GD, Jankowski T, Jarvinen M, Aonghusa CN, Nõges T, et al. 2007. Large-scale climatic signatures in lakes across Europe: a meta-analysis. Glob Change Biol. 13:1314–1326.

Bremigan MT, Soranno PA, Gonzalez MJ, Bunnell DB, Arend KK, Renwick WH, Stein RA, Vanni MJ. 2008. Hydrogeomorphic features mediate the effects of land use/cover on reservoir productivity and food webs. Limnol Oceanogr. 53:1420–1433.

Brylinsky M, Mann KH. 1973. An analysis of factors governing productivity in lakes and reservoirs. Limnol Oceanogr. 18:1–14.

Burford MA, Johnson SA, Cook AJ, Packer TV, Taylor BM, Townsley ER. 2007. Correlations between watershed and reservoir characteristics, and algal blooms in subtropical reservoirs. Water Res. 41:4105–4114.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretical approach. 2nd ed. New York (NY): Springer.

Canfield DE Jr, Hoyer MV. 1988. Regional geology and the chemical and trophic state characteristics of Florida lakes. Lake Res Manage. 4:21–31.

Carneiro FM, Nabout JC, Vieira LCG, Roland F, Bini LM. 2014. Determinants of chlorophyll-*a* concentrations in tropical reservoirs. Hydrobiologia. 740:89–99.

Carpenter SR, Caraco NF, Correll DL, Howarth RW, Sharpley AN, Smith VH. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl. 8:559–568.

[CIESIN] Center for International Earth Science Information Network, Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World (GPW), Version 3.

Chambers PA, Meissner R, Wrona FJ, Rupp H, Guhr H, Seeger J, Culp JM, Brua RB. 2006. Changes in nutrient loading in an agricultural watershed and its effects on water quality and stream biota. Hydrobiologia. 556:399–415.

Chen KS, Lei TC. 2001. Reservoir trophic state evaluation using Landsat TM images. J Am Water Resour As. 37:1321–1334.

Cole JJ, Peierls BL, Caraco NF, Pace ML. 1993. Nitrogen loading of rivers as a human-driven process. New York (NY): Springer.

Cottingham KL, Carpenter SR. 1998. Population, community, and ecosystem variates as ecological indicators: phytoplankton responses to whole-lake enrichment. Ecol Appl. 8:508–530.

D'Arcy P, Carignan R. 1997. Influence of catchment topography on water chemistry in southeastern Quebec Shield lakes. Can J Fish Aquat Sci. 54:2215–2227.

Detenbeck NE, Johnston CA, Niemi GJ. 1993. Wetland effects on lake water quality in the Minneapolis/St. Paul metropolitan area. Landscape Ecol. 8:39–61.

Devito KJ, Creed IF, Rothwell RL, Prepas EE. 2000. Landscape controls on phosphorus loading to boreal lakes: implications for the potential impacts of forest harvesting. Can J Fish Aquat Sci. 57:1977–1984.

Dillon PJ, Kirchner WB. 1975. The effects of geology and land use on the export of phosphorus from watersheds. Water Res. 9:135–148.

Dillon PJ, Rigler FH. 1974. The phosphorus–chlorophyll relationship in lakes. Limnol Oceanogr. 19:767–773.

Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. Ecol Lett. 10:1135–1142.

Gemesi Z, Downing JA, Cruse RM, Anderson PF. 2011. Effects of

watershed configuration and composition on downstream lake water quality. J Environ Qual. 40:507–527.

Gordon DN, McMahon TA, Finlayson BL. 1992. Stream hydrology: an introduction for ecologists. West Sussex (UK): John Wiley & Sons.

Gorman MW, Zimmer KD, Herwig BR, Hanson MA, Wright RG, Vaughn SR, Younk JA. 2014. Relative importance of phosphorus, fish biomass, and watershed land use as drivers of phytoplankton abundance in shallow lakes. Sci Total Environ. 466–447:849–855.

Gruner DS, Smith JE, Seabloom WE, Sandin SA, Ngai JT, Hillbrand H, Harpole WS, Elser JJ, Cleland EE, Bracken MES, et al. 2008. A cross-system synthesis of consumer and nutrient resource control on producer biomass. Ecol Lett. 11:740–755.

Guidford S, Depew D, Ozersky T, Hecky R, Smith R. 2013. Nearshore-offshore differences in planktonic chlorophyll and phytoplankton nutrient status after dreissenid establishment in a large shallow lake. Inland Waters 3:253–268.

Håkanson L. 2005. The importance of lake morphometry and catchment characteristics in limnology—ranking based on statistical analysis. Hydrobiologia. 541:117–137.

Hastie TJ, Tibshirani RJ. 1990. Generalized additive models. New York (NY): Chapman and Hall.

Hecky RE, Smith REH, Barton DR, Guildford SJ, Taylor WD, Charlton MN, Howell T. 2004. The nearshore phosphorus shunt: a consequence of ecosystem engineering by dreissenids in the Laurentian Great Lakes. Can J Fish Aquat Sci. 61:1285–93.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 25:1965–1978.

Huo SL, Ma CZ, He ZS, Xi BD, Su J, Zhang L, Wang J. 2015. Prediction of physico-chemical variables and chlorophyll a criteria for ecoregion lakes using the ratios of land use to lake depth. Environ Earth Sci. 74:3709–3719.

Jones RJ, Bachmann RW. 1977. Trophic status of Iowa lakes in relation to origin and glacial ecology. Hydrobiologia. 57:267–273.

Jones JR, Knowlton MF, Obrecht DV, Cook EA. 2004. Importance of landscape variables and morphology on nutrients in Missouri reservoirs. Can J Fish Aquat Sci. 61:1503–1512.

Jones JR, Obrecht DV, Thorpe AP. 2011. Chlorophyll maxima and chlorophyll: total phosphorus ratios in Missouri reservoirs. Lake Reserv Manage. 27:321–328.

Keitt TH, Bivand R, Pebesma E, Rowlingson B. 2010. rgdal: Bindings for the geospatial data abstraction library. R package version 0.6-25.

Knoll LB, Hagenbuch EJ, Stevens MH, Vanni MJ, Renwick WH, Denlinger JC, Hale RS, Gonzalex MJ. 2015. Predicting eutrophication status in reservoirs at large spatial scales using landscape and morphometric variables. Inland Waters. 5:203–214.

Knoll LB, Vanni MJ, Renwick, WH. 2003. Phytoplankton primary production and photosynthetic parameters in reservoirs along a gradient of watershed land use. Limnol Oceanogr. 48:608–617.

Lehner B, Doll P. 2004. Development and validation of a global database of lakes, reservoirs, and wetlands. J Hydrol. 296:1–22.

Lehner B, Verdin K, Jarvis A. 2008. New global hydrography derived from spaceborne elevation data. Eos Trans AGU. 89:93–94.

Liu WZ, Zhang QF, Liu GH. 2011. Effects of watershed land use and lake morphometry on the trophic state of Chinese lakes: implications for eutrophication control. CLEAN-Soil Air Water. 39:35–42.

Maidment DR. 2002. Arc Hydro: GIS for water resources (Vol. 1). ESRI, Inc.

Markensten H. 2006. Climate effects on early phytoplankton biomass over three decades modified by the morphometry in connected lake basins. Hydrobiologia. 559:319–329.

Marra G, Wood SN. 2011. Practical variable selection for generalized additive models. Comput Stat Data An. 55:2372–2387.

Martin SL, Soranno PA, Bremigan, MT, Cheruvelil KS. 2011. Comparing hydrogeomorphic approaches to lake classification. Environ Manage. 48:957–974.

Meeuwig JJ, Peters RH. 1996. Circumventing phosphorus in lake management: a comparison of chlorophyll *a* predictions from land-use and phosphorus-loading models. Can J Fish Aquat Sci. 53:1795–1806.

Merwade V. 2012. Watershed and stream network delineation using ArcHydro tools. Purdue (IN): University of Purdue, School of Civil Engineering, Printed Lecture Note.

Miranda LE, Andrews CS, Kroger R. 2014. Connectedness of land use, nutrients, primary productions, and fish assemblages in oxbow lakes. Aquat Sci. 76:41–50.

Morin A, Lamoureux W, Busnarda J. 1999. Empirical models predicting primary production from chlorophyll *a* and water temperature for stream periphyton, and lake and ocean phytoplankton. J N Am Benthol Soc. 18:299–307.

[NASA] National Aeronautics and Space Administration's Langley Research Center Atmospheric Science Data Center Surface meteorological and Solar Energy (SSE). 2008. Global horizontal irradiance, 6.0 data set, 22-year monthly and annual average (July 1983–June 2005), one-degree resolution.

Nõges T. 2009. Relationships between morphometry, geographic location, and water quality parameters of European lakes. Hydrobiologia. 633:33–43.

Oksanen L, Fretwell SD, Arruda J, Niemala P. 1981. Exploitation ecosystems in gradients of primary production. Am Nat. 118:240–261.

Postel S, Carpenter S. 2012. Freshwater ecosystem services. In: Daily G, editor. Nature's services: societal dependence on natural ecosystems. Washington (DC): Island Press. p. 195–214.

Potter PN, Ramankutty N, Bennett EM, Donner SD. 2010. Characterizing the spatial patterns of global fertilizer application and manure production. Earth Interact. 14:1–22.

Potter P, Ramankutty N, Bennett EM, Donner SD. 2011. Global fertilizer and manure, version 1: phosphorus fertilizer application. Palisades (NY): NASA Socioeconomic Data and Applications Center (SEDAC).

R Core Team. 2015. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Reed T, Carpenter SR. 2002. Comparisons of P-yield, riparian buffer strips, and land cover in six agricultural watersheds. Ecosystems. 5:568–577.

Sass GZ, Creed IF, Bayley SE, Devito KJ. 2008. Interannual variability in trophic status of shallow lakes on the Boreal Plain: is there a climate signal? Water Resour Res. 44:W08443.

Sayers MJ, Grimm AG, Schuchmann RA, Deines AM, Bunnell DB, Raymer ZB, Rogers MW, Woelmer WM, Bennion DH, Brooks CN, et al. 2015. A new method to generate a high-resolution global distribution map of lake chlorophyll. Int J Remote Sens. 36:1942–1964.

Smith VH. 2003. Eutrophication of freshwater and coastal marine ecosystems: a global problem. Environ Sci Pollut R. 12:126–139.

Smith VH, Schindler DW. 2009. Eutrophication science: where do we go from here? Trends Ecol Evol. 24:201–207.

Soranno PA, Carpenter SR, Lathrop RC. 1997. Internal phosphorus loading in Lake Mendota: response to external loads and weather. Can J Fish Aquat Sci. 54:1883–1893.

Soranno PA, Hubler SL, Carpenter SR, Lathrop RC. 1996. Phosphorus loads to surface waters: a simple model to account for spatial pattern of land use. Ecol Appl. 6:865–878.

Soranno PA, Spence CK, Wagner T, Webster KE. 2015. Effects of land use on lake nutrients: the importance of scale, hydrologic connectivity, and region. PLoS ONE. 10:e0135454.

Soranno PA, Webster KE, Riera JL, Kratz TK, Baron JS, Bukaveckas PA, Kling GW, White DS, Caine N, Lathrop RC, Leavitt PR. 1999. Spatial variation among lakes within landscapes: ecological organization along lake chains. Ecosystems. 2:395–410.

Turcotte R, Fortin JP, Rousseau AN, Massicotte S, Villeneuve, JP. 2001. Determination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network. J Hydrol. 240:225–242.

[USEPA and EC] US Environmental Protection Agency and Environment of Canada. 1995. The Great Lakes: an environmental atlas and resource book. EN40-349/1995E.

Vanni MJ. 2002. Nutrient cycling by animals in freshwater ecosystems. Ann Rev Ecol System. 33:341–370.

Vanni MJ, Arend KK, Bremigan MT, Bunnell DB, Garvey JE, Gonzalez MJ, Renwick WH, Soranno PA, Stein RA. 2005. Linking landscape and food webs: effects of omnivorous fish and watersheds on reservoir ecosystems. Bioscience. 55:155–167.

Vollenweider RA. 1975. Input-output models: with special reference to the phosphorus loading concept in limnology. Aquat Sci. 37:53–84.

Weibel SR. 1969. Urban drainage as a factor in eutrophication. In: Eutrophication: causes, consequences, correctives. Washington (DC): National Academy of Sciences.

Weller CM, Watzin MC, Wang D. 1996. Role of wetlands in reducing phosphorus loading to surface water in eight watersheds in the Lake Champlain Basin. Environ Manage. 20:731–739.

Wetzel RG. 2001. Limnology: lake and river ecosystems. 3rd ed. Houston (TX): Gulf Professional Publishing.

Weyhenmeyer GA. 2001. Warmer winters: are planktonic algal populations in Sweden's largest lakes affected? Ambio. 30:565–571.

Wood SN. 2015. mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. R package version 1.8-5.