

## Predicting eutrophication status in reservoirs at large spatial scales using landscape and morphometric variables

Lesley B. Knoll, Elisabeth J. Hagenbuch, Martin H. Stevens, Michael J. Vanni, William H. Renwick, Jonathan C. Denlinger, R. Scott Hale & María J. González

To cite this article: Lesley B. Knoll, Elisabeth J. Hagenbuch, Martin H. Stevens, Michael J. Vanni, William H. Renwick, Jonathan C. Denlinger, R. Scott Hale & María J. González (2015) Predicting eutrophication status in reservoirs at large spatial scales using landscape and morphometric variables, *Inland Waters*, 5:3, 203-214

To link to this article: <http://dx.doi.org/10.5268/IW-5.3.812>



Published online: 03 Apr 2017.



Submit your article to this journal [↗](#)



Article views: 8



View related articles [↗](#)



View Crossmark data [↗](#)

# Predicting eutrophication status in reservoirs at large spatial scales using landscape and morphometric variables

Lesley B. Knoll,<sup>1\*</sup> Elisabeth J. Hagenbuch,<sup>2†</sup> Martin H. Stevens,<sup>2</sup> Michael J. Vanni,<sup>2</sup> William H. Renwick,<sup>3</sup> Jonathan C. Denlinger,<sup>4</sup> R. Scott Hale,<sup>4</sup> and María J. González<sup>2</sup>

<sup>1</sup> Lacawac Sanctuary Field Station, Lake Ariel, PA, USA

<sup>2</sup> Department of Biology, Miami University, Oxford, OH, USA

<sup>3</sup> Department of Geography, Miami University, Oxford, OH, USA

<sup>4</sup> Ohio Department of Natural Resources, Division of Wildlife, Columbus, OH, USA

\* Corresponding author: knollb@miamioh.edu

† L.B.K. and E.J.H. contributed equally to this work

Received 26 November 2014; accepted 17 April 2015; published 13 May 2015

## Abstract

Aquatic ecosystem management requires knowledge of the links among landscape-level anthropogenic disturbances and aquatic ecosystem properties. With large catchment area to surface area ratios (CA:SA), reservoirs often receive substantial terrestrial subsidies and can be particularly sensitive to eutrophication. Reservoir numbers and attendant management problems are increasing, and tools are needed to categorize their eutrophication status. We analyzed a dataset of 109 reservoirs in Ohio (USA) in an effort to classify eutrophication status using landscape-level features and reservoir morphometry. These predictor variables were selected because they are relatively stable and easily measured. We employed regression tree analysis and used a composite eutrophication variable as our response variable. Our regression tree analysis accurately divided 67% of Ohio reservoirs into 4 eutrophication status groups using 3 predictor variables: percentage of catchment area composed of agriculture versus forest; maximum reservoir depth; and CA:SA. We can infer that reservoirs with catchments containing >71% forest will likely be oligotrophic to mesotrophic. For reservoirs with <71% catchment forest, trophic status is determined by the relative extent of catchment row crops and either CA:SA or maximum depth. We applied our regression tree to a subset of reservoirs in the Environmental Protection Agency's National Lakes Assessment (NLA;  $n = 339$  reservoirs). With a few exceptions, we categorized NLA reservoirs by eutrophication status despite their broad geographical range across the contiguous USA. Our results show that a few easily measured, stable parameters can classify reservoir eutrophication status. Models like ours may be useful for broad-scale management decisions.

**Key words:** catchment, classification, eutrophication, land use, morphometry, reservoirs

## Introduction

For nearly 100 years, limnologists have classified lakes by productivity level and trophic state categories. Lake classification schemes, or indices, were developed using one or more parameters that correlate with lake productivity, often including phytoplankton biomass measured by chlorophyll *a* (Chl-*a*), total phosphorus (TP), total nitrogen (TN), Secchi disk transparency, or some combination of these (Carlson 1977, Vollenweider and Kerekes 1982,

Havens 1994, Nürnberg 1996). These indices have proven to be useful for making policy and supporting lake management decisions. For example, the Organisation for Economic Co-operation and Development (OECD) established trophic state classifications (Vollenweider and Kerekes 1982), and these criteria are used by many international governmental agencies to guide management decisions. Lake managers may not have access to limnological data for all the lakes in their portfolio, however, and therefore need predictive models based on parameters

they can easily obtain. Predictive indices are particularly important because eutrophication of freshwater and marine ecosystems remains one of the most globally prevalent environmental problems, despite decades of awareness, research, and mitigation (Diaz and Rosenberg 2008, Dodds et al. 2009, Smith and Schindler 2009).

In more recent efforts to understand the causes and consequences of eutrophication, relationships between nutrients and primary productivity have been heavily studied (Smith 2003, Schindler 2006, Smith and Schindler 2009). We also now understand that land cover (often agriculture) and lake or catchment morphometric characteristics (e.g., lake maximum depth, hydraulic flushing rates) are strongly related to nutrient concentrations and productivity in natural lakes and constructed reservoirs (Arbuckle and Downing 2001, Prepas et al. 2001, Knoll et al. 2003, Jones et al. 2004, 2008, Bremigan et al. 2008). Here, we focus on reservoirs because the effects of landscape disturbances may be especially pronounced in these waterbodies, which tend to have smaller surface water areas (Whittier et al. 2002) and larger catchment area to surface water area ratios (CA:SA) compared to natural lakes (Kimmel et al. 1990). In addition, the number of reservoirs is increasing worldwide (Downing et al. 2006); they are essential for supplying water for crop irrigation, livestock, drinking water, recreation, and fisheries in many areas, and they are important in sequestering carbon regionally and globally (Cole et al. 2007, Tranvik et al. 2009, Knoll et al. 2013, 2014).

Much knowledge has been gained regarding lake classification schemes and the internal and external controls on trophic status, but the complexity of their interactions and high variability make relationships difficult to predict. Despite this complexity, agencies tasked with managing numerous reservoirs need to make decisions (e.g., where to stock fish, where to manage cyanotoxins) based on limited or no in-lake data. Thus, a recent emphasis is to create predictive models for lake management in which landscape-level attributes are explicitly incorporated and cost-effective approaches are considered (Catherine et al. 2010, Soranno et al. 2010, Cross and Jacobson 2013). These predictive models will be particularly powerful if relationships are robust across broad spatial scales. Non-parametric techniques are an ideal approach to gain insight into the drivers that predict eutrophication levels in reservoirs. We focus on regression tree analysis, a machine learning decision-making tool, because it allows us to use a large dataset and explore complex relationships among predictor variables (De'ath and Fabricius 2000). Although nonparametric techniques such as decision trees are beginning to be used in aquatic ecosystem management, they continue to be underused, even though they are effective (Catherine et al. 2010, Soranno et al. 2010).

The main goal of this study was to develop a model for managing aquatic systems that uses readily available variables to predict eutrophication status in reservoirs. We examined the ability of landscape-level variables (percent of land cover type and CA:SA) and within-reservoir morphometric features (depth and surface area) to predict variables related to eutrophication (chlorophyll, suspended solids, and nutrient concentrations) in reservoirs. Specifically, we developed a regression tree model to predict reservoir eutrophication level using an extensive dataset ( $n = 109$ ) in a specific geographic region in the Midwestern USA (Fig. 1). We then evaluated whether this model could be applied to a much broader scale (contiguous USA) by using the US Environmental Protection Agency's (EPA) 2007 National Lake Assessment (NLA) dataset and selecting reservoirs similar in morphometry to the Ohio reservoirs ( $n = 339$ ). Finally, we compared our regression tree results to a well-known lake trophic status classification scheme (Nürnberg 1996) to examine the predictive performance of our model.

## Methods

### Study area

We sampled 109 reservoirs located throughout Ohio, USA, whose catchments contain a variety of land cover types (Fig. 1; Table S1) and a wide range of eutrophication levels (Table 1). A majority of these reservoirs are tributary reservoirs ( $n = 89$ ) constructed using dams throughout the 1900s, and a smaller number are considered canal and portage reservoirs ( $n = 8$ ) and modified glacial kettle/pothole lakes and quarries ( $n = 12$ ). We included all waterbody types because we aimed to create a model suitable for the entire state. The study reservoirs represent a watershed gradient for forested, agricultural, and developed land use (Fig. 1; Table 1). Wetland cover in the watersheds was minimal with an average of 0.23% and a median of 0.95% cover.

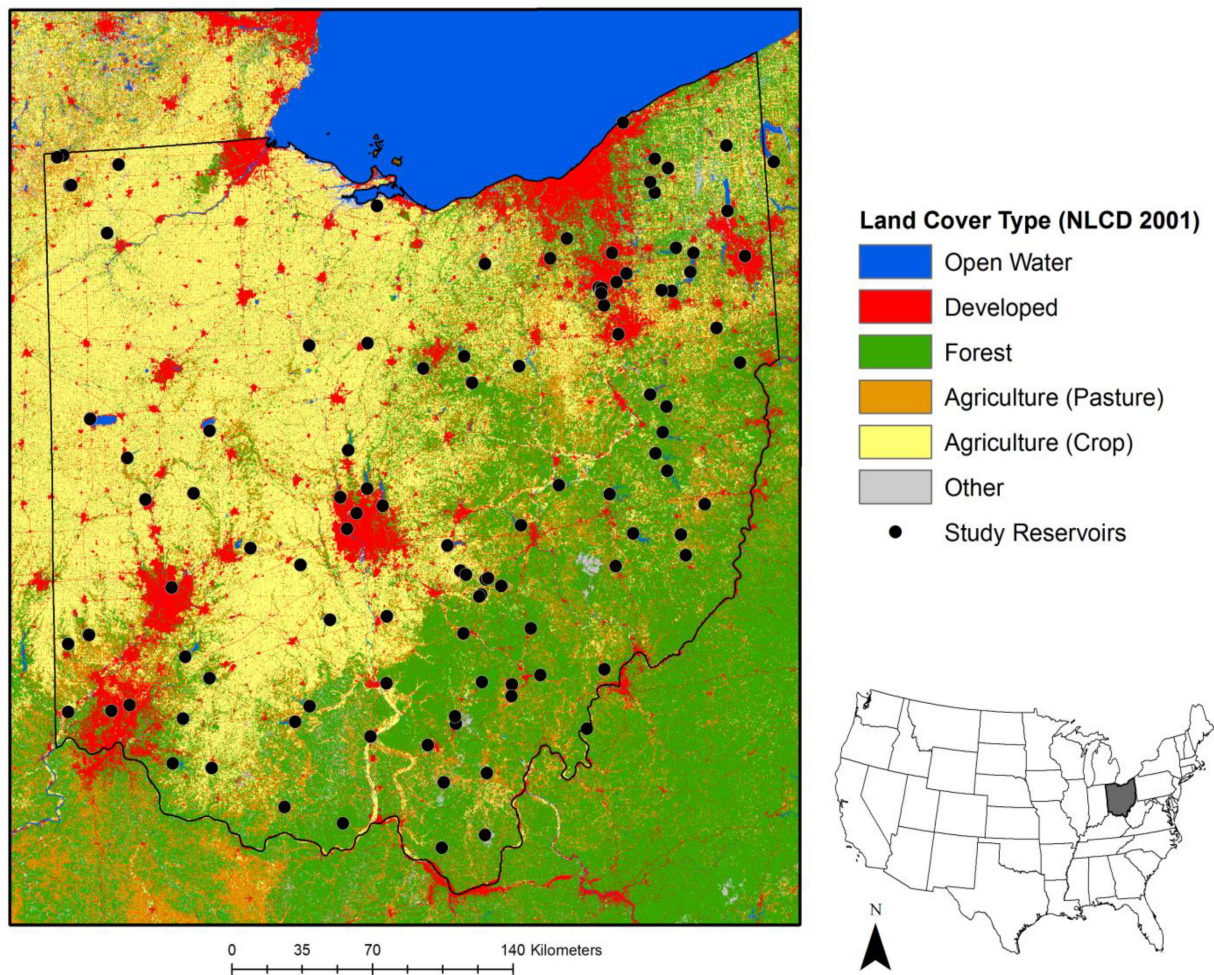
### Sample collection and analyses

All reservoirs were sampled at least once during July or August in 2006 or 2007. We sampled 34 of these reservoirs in both years to assess whether sampling some reservoirs in one year and the others in the next year is likely to provide comparable data. In addition, 10 reservoirs were sampled at least once per month during July and August of both 2006 and 2007.

We collected an integrated sample of epilimnetic water using a flexible plastic tube sampler. The epilimnion was defined as the surface through the deepest depth at which the dissolved oxygen remained  $>2 \text{ mg L}^{-1}$ . Water samples

**Table 1.** Range and mean values of environmental parameters for all Ohio reservoirs and the subset of reservoirs used in 2006 vs. 2007 regressions.

	All Reservoirs (n = 109)		Reservoirs sampled in 2006 and 2007 (n = 34)	
	Range	Mean	Range	Mean
Chlorophyll <i>a</i> ( $\mu\text{g L}^{-1}$ )	1.4–365.1	38.7	3.4–111.86	32.9
Nonvolatile Suspended Solids ( $\text{mg L}^{-1}$ )	0–58.8	3.4	0–6.7	1.0
Total Phosphorus ( $\mu\text{g L}^{-1}$ )	11.9–715.3	76.3	14.3–255.4	51.3
Total Nitrogen ( $\mu\text{g L}^{-1}$ )	129.6–5301.8	1074.4	353.5–4359.6	924.1
Maximum Depth (m)	1.0–15.0	6.3	3.0–16.0	7.2
Reservoir Area ( $\text{km}^2$ )	0.07–65.1	3.9	0.1–65.1	6.6
Catchment Area:Reservoir Surface Area	1.2–1730.0	105.7	1.8–371.2	59.0
Land Cover: % Developed	1.1–86.9	14.6	1.2–65.6	16.6
Land Cover: % Forest	0–94.0	36.3	5.0–83.2	36.0
Land Cover: % Agricultural (Pasture)	0–45.5	12.7	0–30.0	12.7
Land Cover: % Agricultural (Crop)	0.4–78.4	25.0	0–78.4	21.9
Land Cover: % Agricultural (Total)	0–86.2	37.7	0–86.2	34.6

**Fig. 1.** Study locations and land cover using the 2001 National Land Cover Database.



were taken near dam outflows by boat, representing the deepest part of each reservoir. Water samples for Chl-*a* and nonvolatile suspended solids (NVSS) were filtered onto Pall A/E glass fiber filters (1.0  $\mu\text{m}$  nominal pore size, preweighed for NVSS) and frozen for later laboratory analysis. NVSS concentration was used as a quantitative index of inorganic turbidity (Knowlton and Jones 2000). Although NVSS concentration can increase due to wind-induced sediment resuspension in shallow areas of a reservoir, it also increases greatly after large storms due to soil erosion and runoff (Vanni et al. 2006). NVSS therefore provides a potentially useful indicator of catchment influence (Jones and Knowlton 2005), particularly in deeper areas where sediment resuspension has less influence. It has also been suggested that parameters of nonalgal turbidity are appropriate to consider for trophic status in reservoirs that receive high amounts of catchment-derived sediment because Chl-*a*–water clarity relationships differ in these reservoirs when compared to natural lakes (Walker 1984). Water samples for TP and TN were acidified and stored at 4 °C prior to analysis.

Water sample processing for Chl-*a*, NVSS, and TP followed the analytical procedures outlined in Knoll et al. (2003). Briefly, Chl-*a* was extracted from filters in the dark at 4 °C using acetone and measured on a Turner model TD-700 fluorometer (Sunnyvale, CA, USA). To quantify NVSS, filters were dried (24 h at 60 °C), weighed, ashed to remove organic matter (4 h at 550 °C), and then reweighed. Unfiltered water samples were assayed for TP and TN. TP was analyzed with a Lachat QC 8000 FIA autoanalyzer (Lachat Instruments, Loveland, CO, USA) using the acid molybdate method, following digestion with potassium persulfate. TN was converted to nitrate via low-N potassium persulfate digestion and analyzed using second-derivative spectroscopy on a Perkin Elmer (Waltham, MA, USA) Lambda 35 UV/VIS spectrophotometer (Crumpton et al. 1992).

### Reservoir morphometric parameters

The maximum sampling depth used for dissolved oxygen/temperature profiles (MaxDepth) was used as a surrogate for the maximum physical depth because detailed bathymetric data were not available for all reservoirs. Available bathymetric maps were used to compare actual maximum depth to the maximum sampling depth and were positively correlated ( $r^2 = 0.60$ ,  $p < 0.001$ ,  $n = 82$ ). Reservoir surface areas (Area) were calculated in ESRI ArcGIS 9.3 (ESRI, Inc., Redlands, CA, USA).

### Landscape parameters

All GIS data were quantified in ESRI ArcGIS 9.3. A 30 m digital elevation model (DEM) of Ohio was downloaded

from the United States Geological Survey national map seamless server (<http://seamless.usgs.gov/>). Reservoir catchments were delineated from reservoir polygon shapefiles using the Arc Hydro Tools 9 extension. DEM manipulation steps included DEM reconditioning using the National Hydrography Dataset 1:24,000 scale flowlines (<http://nhd.usgs.gov/>) and fill sinks. Land cover percentages for agricultural cropland (PercentAgCrop); agricultural pasture (PercentAgPasture); deciduous, evergreen, and mixed forest (PercentForest); and open, low, medium, and high intensity developed (PercentDeveloped) were calculated using the designations provided by the 30 m resolution 2001 National Land Cover Database (NLCD; Level II overall accuracy =  $64\% \pm 2$ ; Wickham et al. 2004). The NLCD classifies land cover into 16 classes using Landsat satellite data. We were unable to ground truth watershed land cover classifications. A new NLCD was released in 2011 for 2006 land cover data, but for the state of Ohio, the percentage change in land cover from 2001 to 2006 was  $<1\%$  (Fry et al. 2011); therefore, we do not expect that using the newer NLCD would change our study results. CA:SA was calculated using the delineated catchments and digitized reservoir surface areas.

### Relationships among environmental parameters in reservoirs

We compared eutrophication indicators (Chl-*a*, TP, TN, and NVSS) in the subset of reservoirs sampled both years by regressing 2006 vs. 2007 values to determine the efficacy of sampling in just 1 year (which was the case for the majority of the reservoirs). In addition, we conducted simple linear regressions between pairs of parameters (Chl-*a*, TP, TN, and NVSS) as well as relationships between each eutrophication indicator and morphometric (MaxDepth, Area) and landscape-level parameters (PercentAgCrop, PercentAgPasture, PercentForest, PercentDeveloped, CA:SA) for each year separately to verify similarity between year-to-year patterns. Based on similarities between 2006 and 2007 regression results, we combined these datasets and conducted regressions as above using the full dataset. When multiple samples were available for a reservoir within 1 year (10 reservoirs), averages were time-weighted using the proportion of time represented by each sample. Finally, for reservoirs sampled in both years, values were averaged to obtain a single value for each reservoir; thus, all reservoirs were given equal weight regardless of sampling effort. All statistical analyses were performed in R (R Development Core Team 2009). To improve normality, all data were  $\log_{10}(x+1)$  transformed with the exception of land cover percentage, which was arcsine square-root transformed (Knoll et al. 2003, Babler et al. 2011).

A regression tree analysis was performed using the *rpart* package in R (Therneau and Atkinson 2008) to assess whether we could use landscape and morphometric variables to accurately group reservoirs according to eutrophication indicators. Because our dataset consisted of continuous variables, the “anova” method was selected. This method divides the reservoirs by minimizing the residual sums of squares within each group at each level of the tree, with each split decision being made independently of prior splits (Maindonald and Braun 2007). Prior to implementing the regression tree analysis, eutrophication indicators (Chl-*a*, TP, TN, and NVSS) were entered into a principal components analysis (PCA) to create a composite “eutrophication” variable, which was used as the response variable in the regression tree analysis. We chose this approach because of high levels of collinearity among eutrophication indicators (Chl-*a*, NVSS, TP, and TN). Possible predictor variables entered into the analysis were MaxDepth, Area, PercentAgCrop, PercentForest, and CA:SA. A regression tree was over-fitted and then trimmed using the one-standard-deviation rule to select the minimum size tree, where the cross-validated error is less than the minimum cross-validated error plus one standard deviation. When considering the accuracy of the regression tree, both relative error and cross-validated error were considered. The relative error provides an error estimate based on the current dataset from which the current regression tree was built, while the cross-validated error uses subsets of the data to estimate the accuracy of the model for new data (Maindonald and Braun 2007).

We compared our eutrophication level groups resulting from the regression tree to a widely used lake classification scheme (Nürnberg 1996). We also calculated reference conditions for Chl-*a*, TP, and TN for the Ohio reservoir dataset as an additional check on the trophic state categories we assigned. We used a recent regression technique with watershed land use classes with anthropogenic impacts (e.g., agricultural, developed) as the predictor variable(s) and eutrophication parameters (Chl-*a*, TP, or TN) as the response variable (Dodds and Oakes 2004, Dodds et al. 2006). The y-intercept of these regressions was used to estimate the reference values (i.e., in the absence of disturbed watershed land use).

### Applying the Ohio reservoir regression tree to EPA National Lakes Assessment reservoirs

We used the US EPA 2007 NLA database (see [http://water.epa.gov/type/lakes/NLA\\_data.cfm](http://water.epa.gov/type/lakes/NLA_data.cfm) for more information about NLA) to assess how well the regression tree developed for the Ohio reservoirs can predict eutrophication status in other reservoirs. Briefly, the NLA dataset includes natural and artificial lakes >10 ac (4 ha) in

surface area and at least 1 m deep, selected randomly by the EPA to represent waterbodies found in the contiguous USA (i.e., excluding Hawaii and Alaska). For our study, we selected only constructed reservoirs within the NLA dataset that fell within the range of Ohio reservoirs for each of the following parameters: reservoir surface area, catchment area, maximum depth, and CA:SA (Table 1). We restricted the NLA dataset to reservoirs sampled in July and August to match the Ohio reservoir dataset. As with the Ohio data, in the cases where a reservoir was sampled in both July and August, we took the mean of the variable of interest. Based on these criteria, we used 339 NLA reservoirs. We applied the tree developed using Ohio reservoirs to the NLA reservoirs to test the effectiveness of our regression tree model. Specifically, we categorized the NLA reservoirs into 5 groups using the tree splits based on landscape and morphometric values (details below).

## Results

All correlations between 2006 versus 2007 water quality parameters were significant ( $r^2 = 0.23\text{--}0.86$ ;  $p < 0.004\text{--}0.001$ ), with the strongest relationships (highest  $r^2$ ) observed for TP and Chl-*a* ( $r^2 = 0.86$  and  $0.55$ , respectively). With the exception of NVSS, 95% confidence intervals for slopes and intercepts of all regression lines were inclusive of 1 and 0, respectively. As anticipated, eutrophication indicators were highly correlated with each other (Table 2). Strong positive relationships were observed among Chl-*a*, TP, and TN. Additionally, NVSS was positively correlated with Chl-*a*, TN, and TP.

We also detected significant relationships between eutrophication indicators and morphometric parameters (Table 2). MaxDepth was negatively related to Chl-*a*, TP, TN, and NVSS. Reservoir area showed only a weak positive relationship with Chl-*a*. Of the landscape parameters, CA:SA exhibited weak positive relationships with Chl-*a*, NVSS, TP, and TN. We detected significant positive relationships between PercentAgCrop and Chl-*a*, NVSS, TP, and TN, while PercentForest was negatively related to these parameters (note that PercentAgCrop and PercentForest are strongly negatively related). PercentDeveloped showed only a weak positive relationship with TP.

The first principal component of the eutrophication indicators (Chl-*a*, NVSS, TP, and TN) explained 71% of the variance among reservoirs and was significantly correlated with all water quality parameters (Pearson correlations: Chl-*a*  $-0.88$ , NVSS  $-0.68$ , TP  $-0.94$ , TN  $-0.84$ ;  $p < 0.001$ ). The scores from the first principal component axis for individual reservoirs representing this composite eutrophication variable ranged from  $-4.65$  (high eutrophication)

**Table 2.** Values of  $r^2$  for simple linear regressions among eutrophication indicators and between single eutrophication indicators and morphometric and landscape parameters (PercentDeveloped: open, low, medium, and high intensity developed; PercentForest: deciduous, evergreen, and mixed forest; PercentAgCrop: agricultural row crop; PercentAgPasture: agricultural pasture; ns = not significant, (–) and (+) refer to the direction of the relationship, \* $p < 0.05$ , \*\* $p < 0.001$ ).

	Chlorophyll <i>a</i> ( $\mu\text{g L}^{-1}$ )	Nonvolatile suspended solids ( $\text{mg L}^{-1}$ )	Total phosphorus ( $\mu\text{g L}^{-1}$ )	Total nitrogen ( $\mu\text{g L}^{-1}$ )
<b>Water Quality</b>				
Nonvolatile suspended solids ( $\text{mg L}^{-1}$ )	0.14 (+)**	—	—	—
Total phosphorus ( $\mu\text{g L}^{-1}$ )	0.65 (+)**	0.41 (+)**	—	—
Total nitrogen ( $\mu\text{g L}^{-1}$ )	0.50 (+)**	0.14 (+)**	0.50 (+)**	—
<b>Morphometric</b>				
Reservoir surface area ( $\text{m}^2$ )	0.10 (+)**	ns	ns	ns
Maximum sampling depth (m)	0.19 (–)**	0.23 (–)**	0.34 (–)**	0.10 (–)**
<b>Landscape</b>				
PercentDeveloped	ns	ns	0.04 (+)*	ns
PercentForest	0.11 (–)**	0.08 (–)*	0.24 (–)**	0.43 (–)**
PercentAgCrop	0.29 (+)**	0.18 (+)**	0.31 (+)**	0.35 (+)**
PercentAgPasture	ns	ns	ns	ns
Catchment area:reservoir surface area ratio	0.08 (+)*	0.10 (+)**	0.11 (+)**	0.06 (+)*

cation levels) to 3.21 (low eutrophication levels) and were used as the response variable in the regression tree analysis. Of the 5 morphometric and landscape parameters entered as possible predictors in the regression tree analysis (MaxDepth, Area, PercentAgCrop, PercentForest, and CA:SA), all but Area were included in the final model, and these remaining 4 parameters explained 67% of the variation among reservoirs (relative error = 0.33, where relative error =  $1 - r^2$ ). The cross-validated error rate (0.48) showed that after resubstitution using randomly selected subsets of the data, the model explained 52% of the variance among reservoirs.

The best regression tree model divided the reservoirs into 5 groups with mean eutrophication (PCA) scores of  $-2.68$ ,  $-0.833$ ,  $-0.707$ ,  $0.634$ , and  $2.09$  (Fig. 2); however, each split decision was made independently of prior splits, and the second and third groups ( $-0.833$  and  $-0.707$ ) showed similar values in terms of eutrophication level. Therefore, reservoirs were ultimately classified into 4 groups (“very high,” “high,” “moderate,” and “low”) according to the composite eutrophication indicator (Fig. 2).

The first “branch” or split of the tree separated reservoirs based on land use, specifically PercentAgCrop, whereas the next 2 divisions were based on MaxDepth or PercentForest (Fig. 2). CA:SA was also an important variable. Thus, very high eutrophication levels (i.e., severe symptoms of eutrophication) were associated with shallow reservoirs in catchments containing relatively high levels of row crop cover. The high eutrophication level group

contained reservoirs that were either with (a) relatively deep lakes in catchments with high row crop agriculture, or (b) lakes located in catchments with lower row crop cover and large CA:SA. In terms of predictor variables, the moderate eutrophication level group differed from the high eutrophication level group by having smaller CA:SA. As expected, the low eutrophication level group had low levels of row crop cover.

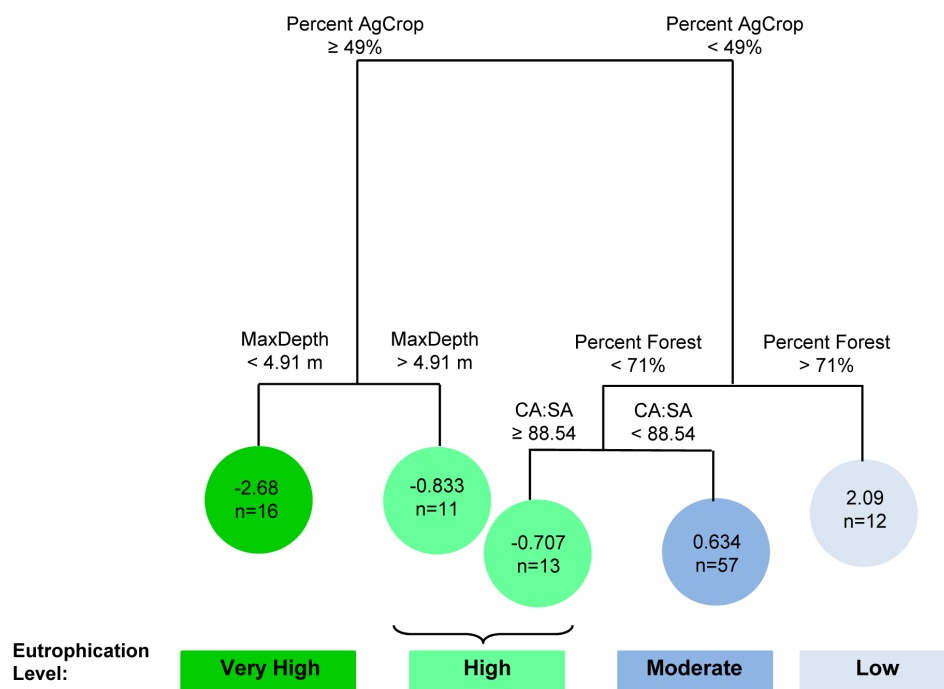
Using the regression tree developed with the Ohio reservoirs and the associated branch splits based on land cover, reservoir depth, and CA:SA, we classified the NLA reservoirs as low, moderate, high, or very high. The NLA dataset was dominated by moderate and high category reservoirs, with few in the low category and even fewer in the very high category (Table 3). With exceptions, particularly for TN (which tended to be higher in Ohio reservoirs), median concentrations for the regression tree eutrophication levels were generally similar between the Ohio and NLA datasets (Table 3; Fig. 3).

Our 4 eutrophication level groups (i.e., low to very high) correspond roughly to 4 groups derived from a widely used lake classification scheme (Nürnberg 1996): oligotrophic, mesotrophic, eutrophic, and hypereutrophic categories based on group median Chl-*a*, TP, and TN concentrations for both the Ohio and NLA datasets (Table 3). Differences between the regression tree groupings and trophic state levels did emerge with our regression tree, however, both with the Ohio and NLA datasets. The median TP concentration in our low category is classified

**Table 3.** Summary of median Chl-*a*, TN, and TP concentrations, the number (n) and percentage of reservoirs in each regression tree eutrophication level, and the trophic state level as categorized by the median concentration for each limnological variable (Nürnberg 1996). \* The Ohio and NLA datasets were classified into eutrophication levels by the regression tree (i.e., low, moderate, high, or very high).

Reservoir dataset (total # reservoirs)	Regression tree eutrophication level	n	% of reservoirs	Median Chl- <i>a</i> ( $\mu\text{g L}^{-1}$ )	Trophic state*	Median TP ( $\mu\text{g L}^{-1}$ )	Trophic state*	Median TN ( $\mu\text{g L}^{-1}$ )	Trophic state*
Ohio (109)	Low	12	11	7.9	mesotrophic	20	mesotrophic	334	oligotrophic
	Moderate	58	52	20	eutrophic	37	eutrophic	677	eutrophic
	High	24	22	38	hypereutrophic	73	eutrophic	1000	eutrophic
	Very High	16	15	73	hypereutrophic	188	hypereutrophic	1580	hypereutrophic
NLA (339)	Low	61	18	4	mesotrophic	13	mesotrophic	274	oligotrophic
	Moderate	175	52	7.8	mesotrophic	24	mesotrophic	504	mesotrophic
	High	93	27	15	eutrophic	56	eutrophic	704	eutrophic
	Very High	10	3	83	hypereutrophic	329	hypereutrophic	2234	hypereutrophic

\*Trophic status categories (Nürnberg 1996): Chl-*a* (oligotrophic =  $<3.5$ , mesotrophic =  $3.5\text{--}9$ , eutrophic =  $9.1\text{--}25$ , hypereutrophic =  $>25 \mu\text{g L}^{-1}$ ), TP (oligotrophic =  $<10$ , mesotrophic =  $10\text{--}30$ , eutrophic =  $31\text{--}100$ , hypereutrophic =  $>100 \mu\text{g L}^{-1}$ ), and TN oligotrophic =  $<350$ , mesotrophic =  $350\text{--}650$ , eutrophic =  $651\text{--}1200$ , hypereutrophic =  $>1200 \mu\text{g L}^{-1}$ )



**Fig. 2.** Regression tree analysis showing significant predictors of a composite response variable representing eutrophication levels. The composite response variable was created from a PCA using Chl-*a*, TP, TN, and NVSS. Low values indicate high eutrophication levels and high values indicate low eutrophication levels. Although the tree divided the reservoirs into 5 groups, 2 groups had similar eutrophication levels; therefore, 4 eutrophication levels were designated. The height of the branch coincides with relative reduction on the total sums of squares. Abbreviations are as defined in Table 2.



as mesotrophic by Nürnberg's (1996) classification, and TP concentrations in our moderate category were mesotrophic or eutrophic as classified by Nürnberg (1996). Similar results were found for Chl-*a*. Only for TN were reservoirs in the low level classified as oligotrophic by Nürnberg's criteria.

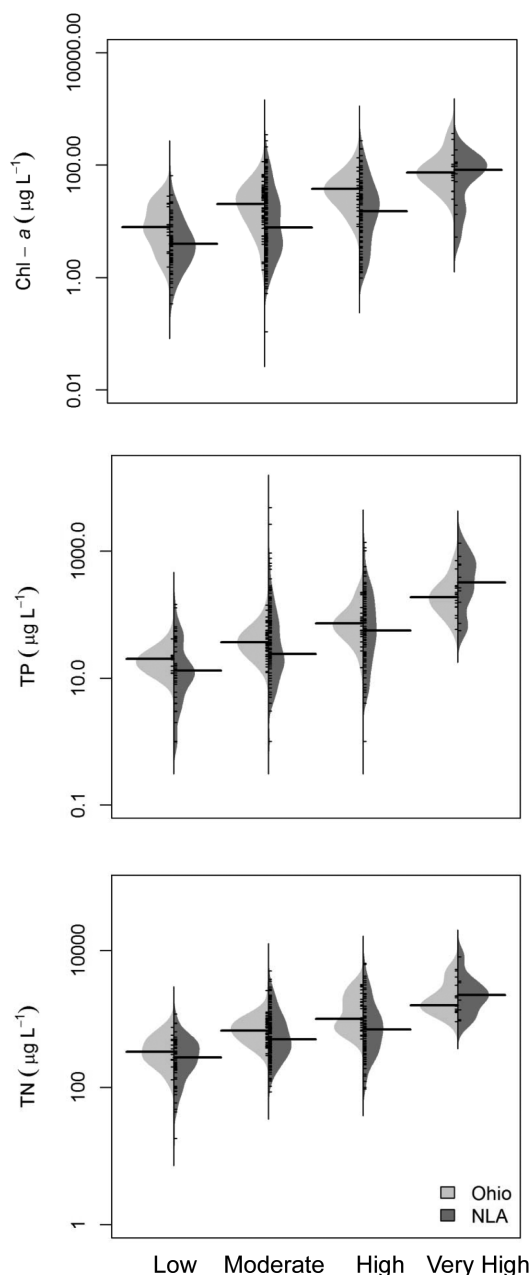


Figure 3. Asymmetric beanplots showing the range and median concentrations within each eutrophication level group (as classified by regression tree) for the Ohio (light gray; left side of beanplot) and NLA (dark gray; right side of beanplot) datasets. Individual reservoir observations are shown with short black lines and the medians with long black lines. The shape of the bean depicts the distribution of the data.

Using a regression technique for determining lake and reservoir reference conditions (Dodds et al. 2006), we found reference conditions in Ohio reservoirs to be 28  $\mu\text{g L}^{-1}$  for TP, 341  $\mu\text{g L}^{-1}$  for TN, and 7.1  $\mu\text{g L}^{-1}$  for Chl-*a*.

## Discussion

Our results demonstrate that a model developed with relatively easily obtained variables can be used to predict eutrophication status in reservoirs. Notably, we effectively implemented a regression tree model built on reservoirs from a specific geographic region in Midwestern USA to classify NLA reservoirs located across the contiguous USA. Our regression tree provides resource managers with a tool to quickly identify the eutrophic condition of reservoirs based on land cover and morphometry. For example, using the Ohio dataset, we found that very high eutrophication status reservoirs were only found in catchments composed of at least 49% row crop agriculture, which generally agrees with a recent effort in Minnesota examining landscape influences on lake TP concentrations (Cross and Jacobson 2013). The authors found a curvilinear relationship with catchment disturbance on TP; catchments with at least 40% disturbance had much higher lake TP concentrations than those with less disturbance. Overall, our results suggest that eutrophication in reservoirs across Ohio and the contiguous US is relatively common (Table 3). In both the Ohio and NLA datasets, only a small proportion of reservoirs were considered oligotrophic, and a majority was categorized as mesotrophic to hypereutrophic. This supports decades of research highlighting cultural eutrophication in freshwaters (Smith et al. 2014).

Our regression tree model, which used a composite eutrophication response variable, offers a unique approach for predicting eutrophication status in reservoirs. The composite variable was created from a PCA analyses and included Chl-*a*, TP, TN, and NVSS. Many studies focused on predicting eutrophication status include Chl-*a*, TP, and TN as response variables, but it is not as common to include NVSS or another metric specifically measuring nonalgal turbidity. Reservoir scientists have long recognized that Chl-*a* and water clarity relationships are weak in reservoirs with high sediment turbidity (Jones and Bachmann 1978, Lind 1986, Jones and Knowlton 1993), and that a measure of nonalgal turbidity is an important factor to consider for trophic state classification in this type of reservoir (Walker 1984, Dzialowski et al. 2011). Thus, our regression tree model is one of the first to consider a complete suite of indicators known to be important in reservoir ecosystems.

The extent to which we could more accurately classify reservoirs was likely limited by the unavailability of

additional within-lake and catchment-scale factors potentially influencing reservoir eutrophication level. Important considerations include nutrient cycling within the reservoir, presence of invasive species (e.g., zebra and quagga mussels), variation in agricultural practices within a type of land cover (e.g., tillage practices), ecoregion units (Soranno et al. 2010, Wagner et al. 2011, Cheruvilil et al. 2013), or the distance between a particular type of land cover and the reservoir (King et al. 2007). For example, conservation tillage can result in lower levels of soluble reactive phosphorus and suspended sediments in streams and may thus decrease reservoir eutrophication levels, even though land cover (i.e., percent row crops) is the same as under conventional tillage (Richards and Baker 2002, Renwick et al. 2008). Widespread data on tillage practices are generally not available on a catchment-specific basis, however, and so could not be factored into this study. Where available, the inclusion of detailed information on agricultural practices will most likely aid in explaining relationships between land cover type and eutrophication levels in waterbodies.

Studies of some areas have shown that including the distance between land cover types and a waterbody can improve the relationship between nutrient concentrations and land cover types within a catchment (King et al. 2007, Fraterrigo and Downing 2008). This was not the case for a subset of reservoirs within in our area, where weighting agricultural cover closer to the reservoir more heavily did not improve the relationship obtained using the flat percentage (Hagenbuch 2010). Recent stream analyses suggests that downstream water chemistry is better explained by watershed cover and riparian land use for small, first-order upland streams than for these factors near the downstream sampling sites (Dodds and Oakes 2008). Further, recent work suggests that considering ecological drainage units (ecoregions) allows a better understanding of lake water quality dynamics, including lake productivity (Soranno et al. 2010, Wagner et al. 2011). We did not take ecoregions into account with our Ohio dataset because the geographic region is limited, our reservoirs are not evenly distributed within the units, and there are distinct land cover regions in the state (Fig. 1), so any potentially differences explained by ecoregions would likely also be explained by land cover.

The strongest links between reservoir eutrophication level and land cover type were identified for the 2 primary land cover types within our reservoir dataset: forest and agricultural row crop, which are strongly negatively related to each other. The apparent lack of a significant effect of other types of land cover, such as developed or agricultural pasture land cover, on reservoir eutrophication level may be because inference is limited to the dominant land cover types within the study area. For example, in our Ohio

dataset, the limited number of reservoirs with significant developed land cover renders the relationship with eutrophication level weak. Only 14% of our catchments contained >25% developed land cover, while 39% of the catchments contained at least 25% agricultural row crop, and 58% contained at least 25% forest cover. The NLA dataset also had a low frequency of reservoirs with catchments composed of >25% developed land (10%), while 75% had either >25% agricultural row crop (18%) or forest cover (57%). A recent study took a similar approach as our study for 50 constructed waterbodies in France and found that Chl-*a* measurements were best predicted by a combination of land use and morphometric parameters (Catherine et al. 2010). Unlike our study, many of the reservoirs in the Catherine et al. (2010) study were dominated by urban land use. In combination, these 2 studies suggest that consideration of both land use and morphometry allows accurate prediction of trophic status in reservoirs, regardless of the dominant land use type.

Many trophic state indices have been developed with natural lakes or a mix of natural and constructed systems, but there are many physical and functional differences between reservoirs and natural lakes, such as morphometry, water residence time, and CA:SA (Kimmel et al. 1990). Nonetheless, we compared our regression tree eutrophication levels (i.e., low to very high) to trophic state levels to better understand the utility of our model. We were able to successfully break our data into 4 groups of differing eutrophication levels that approximately match established trophic state levels (Nürnberg 1996), which suggests that our regression tree is effective at categorizing reservoirs using landscape-level and morphometric characteristics. However, the regression tree output only completely matched Nürnberg's scheme (i.e., low = oligotrophic, very high = hypereutrophic, etc.) for TN with the NLA reservoir dataset, possibly because (1) Nürnberg's dataset is primarily composed of natural lakes, which are potentially more likely to have oligotrophic systems than constructed reservoirs, (2) Nürnberg's scheme was developed using both a global and North American set of lakes, and (3) Nürnberg's scheme used a regression approach to verify earlier trophic state categories based on TP. The global set included lakes from 3 continents, and the North American lakes were located in Central Ontario, Southern Ontario, Quebec, and eastern USA. In both datasets, the lakes ranged from hardwater to softwater and from oligotrophic to hypereutrophic; however, the Ohio reservoirs had a much larger range for Chl-*a*, TP, and TN with much higher maximum values (5–9 times) than Nürnberg's North America dataset. The Ohio reservoirs fell within the global lake range for TP and TN but not for Chl-*a* (maximum Chl-*a* in Nürnberg global dataset = 189  $\mu\text{g L}^{-1}$  and in Ohio dataset = 365  $\mu\text{g L}^{-1}$ ).

In support of our assigned categories, we found reference TP conditions in the Ohio reservoirs to be similar to a study in Kansas lakes and reservoirs (Dodds et al. 2006), and the Ohio reference value is slightly below  $30 \mu\text{g L}^{-1}$ , Nürnberg's (1996) value for transitioning from mesotrophy to eutrophy. Our TP reference value is also below the threshold of  $30 \mu\text{g L}^{-1}$  generally considered to indicate an increased probability of cyanobacteria blooms in freshwaters (Downing et al. 2001). Our TN reference value is just below the  $350 \mu\text{g L}^{-1}$  cut-off from oligotrophy to mesotrophy with Nürnberg's (1996) criteria. Finally, our Chl-*a* reference is below the mesotrophic to eutrophic transition at  $9 \mu\text{g L}^{-1}$ . Taken together, this information suggests that in the absence of degraded land use (i.e., reference conditions), Ohio reservoirs would fall into the oligotrophic to mesotrophic range and also within the ranges of Nürnberg's classification scheme.

In summary, we developed a predictive classification model using landscape-level and morphometric parameters to predict reservoir eutrophication status. The regression tree can also be extended to broader geographic regions as was shown with the NLA dataset. There was some variability with regression tree predictions and a common trophic state index, particularly between the oligotrophic–mesotrophic boundary, but the model generally performed well with both reservoir datasets. One of the most important implications of our regression tree analysis is that we predicted reservoir trophic status using the relatively easily measured variables of land cover proportions, reservoir depth, and the catchment:reservoir surface area ratio. These parameters are relatively constant from year-to-year (aside from anthropogenic land use change), suggesting that reservoirs can be classified into trophic state without intense in-reservoir sampling. This information should be valuable to stakeholders interested in predicting, assessing, and managing water quality at a relatively low cost. Further, managers may be able to use the model to infer which reservoirs may be at risk of becoming eutrophic and may benefit the most from watershed management.

## Acknowledgements

Funding for this research was provided by the Federal Aid in Sport Fish Restoration Program (F-69-P, Fish Management in Ohio) administered jointly by the US Fish and Wildlife Service and the Ohio Department of Natural Resources Division of Wildlife to M.J. González and M.J. Vanni; National Science Foundation (DEB: 0235755 and DEB: 0743192) to M.J. Vanni, M.J. González, and W.H. Renwick; and a Summer Workshop from Miami University to E.J. Hagenbuch. We would like to thank A.

Bowling, C. Glaholt, and P. Levi for their help during sampling; R. Abbitt and S. Xanakis for their help with GIS; and J. Conroy, the González and Vanni labs, and 3 anonymous reviewers for their valuable comments on the manuscript.

## References

- Arbuckle KE, Downing JA. 2001. The influence of watershed land use on lake N:P in a predominantly agricultural landscape. *Limnol Oceanogr.* 46:970–975.
- Babler AL, Pilati A, Vanni MJ. 2011. Terrestrial support of detritivorous fish populations decreases with watershed size. *Ecosphere.* 2:23.
- 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.
- Carlson RE. 1977. Trophic state index for lakes. *Limnol Oceanogr.* 22:361–369.
- Catherine A, Mouillot D, Escoffier N, Bernard C, Troussellier M. 2010. Cost effective prediction of the eutrophication status of lakes and reservoirs. *Freshwater Biol.* 55:2425–2435.
- Cheruvelil KS, Soranno PA, Webster KE, Bremigan MT. 2013. Multi-scaled drivers of ecosystem state: quantifying the importance of the regional spatial scale. *Ecol Appl.* 23:1603–1618.
- Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ, et al. 2007. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems.* 10:171–184.
- Cross TK, Jacobson PC. 2013. Landscape factors influencing lake phosphorus concentrations across Minnesota. *Lake Reserv Manage.* 29:1–12.
- Crumpton WG, Isenhardt TM, Mitchell PD. 1992. Nitrate and organic N analyses with 2nd-derivative spectroscopy. *Limnol Oceanogr.* 37:907–913.
- De'ath G, Fabricius KE. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology.* 81:3178–3192.
- Diaz RJ, Rosenberg R. 2008. Spreading dead zones and consequences for marine ecosystems. *Science.* 321:926–929.
- Dodds WK, Bouska WW, Eitzmann JL, Pilger TJ, Pitts KL, Riley AJ, Schloesser JT, Thornbrugh DJ. 2009. Eutrophication of US freshwaters: analysis of potential economic damages. *Environ Sci Technol.* 43:12–19.
- Dodds WK, Carney E, Angelo RT. 2006. Determining ecoregional reference conditions for nutrients, secchi depth, and chlorophyll *a* in Kansas lakes and reservoirs. *Lake Reserv Manage.* 22:151–159.
- Dodds WK, Oakes RM. 2004. A technique for establishing reference nutrient concentrations across watershed affected by humans. *Limnol Oceanogr: Methods.* 2:333–341.
- Dodds WK, Oakes RM. 2008. Headwater influences on downstream water quality. *Environ Manage.* 41:367–377.
- Downing JA, Prairie YT, Cole JJ, Duarte CM, Tranvik LJ, Striegl RG,

- McDowell WH, Kortelainen P, Caraco NF, Melack JM, et al. 2006. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol Oceanogr.* 51:2388–2397.
- Downing JA, Watson SB, McCauley E. 2001. Predicting cyanobacteria dominance in lakes. *Can J Fish Aquat Sci.* 58:1905–1908.
- Dzialowski AR, Smith VH, Wang SH, Martin MC, deNoyelles F. 2011. Effects of non-algal turbidity on cyanobacterial biomass in seven turbid Kansas reservoirs. *Lake Reserv Manage.* 27:6–14.
- Fraterrigo JM, Downing JA. 2008. The influence of land use on lake nutrients varies with watershed transport capacity. *Ecosystems.* 11:1021–1034.
- Fry JA, Xian G, Jin S, Dewitz JA, Homer CG, Yang L, Barnes CA, Herold ND, Wickham JD. 2011. National land cover database for the conterminous United States. *Photogramm Eng Remote Sens.* 77:859–864.
- Hagenbuch EJ. 2010. Complex relationships among watershed land cover and reservoir morphometry, productivity, and zooplankton communities [master's thesis]. [Oxford (OH)]: Miami University.
- Havens KE. 1994. Seasonal and spatial variation in nutrient limitation in a shallow subtropical lake (Lake Okeechobee, Florida) as evidenced by trophic state index deviations. *Arch Hydrobiol.* 131:39–53.
- Jones JR, Bachmann RW. 1978. Trophic status of Iowa lakes in relation to origin and glacial geology. *Hydrobiologia.* 57:267–273.
- Jones JR, Knowlton MF. 1993. Limnology of Missouri reservoirs: an analysis of regional patterns. *Lake Reserv Manage.* 8:17–30.
- 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, Knowlton MF. 2005. Suspended solids in Missouri reservoirs in relation to catchment features and internal processes. *Water Res.* 39:3629–3635.
- Jones JR, Knowlton MF, Obrecht DV. 2008. Role of land cover and hydrology in determining nutrients in mid-continent reservoirs: implications for nutrient criteria and management. *Lake Reserv Manage.* 24:1–9.
- Kimmel BL, Lind OT, Paulson LJ. 1990. Reservoir primary production. In: Thornton KW, Kimmel BL, Payne FE, editors. *Reservoir limnology: ecological perspectives*. New York: Wiley Interscience. p. 133–194.
- King RS, Deluca WV, Whigham DF, Marra PP. 2007. Threshold effects of coastal urbanization on *Phragmites australis* (common reed) abundance and foliar nitrogen in Chesapeake Bay. *Estuar Coasts.* 30:469–481.
- 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.
- Knoll LB, Vanni MJ, Renwick WH, Dittman EK, Gephart JA. 2013. Temperate reservoirs are large carbon sinks and small CO<sub>2</sub> sources: results from high-resolution carbon budgets. *Glob Biogeochem Cy.* 27:52–64.
- Knoll LB, Vanni MJ, Renwick WH, Kollie S. 2014. Burial rates and stoichiometry of sedimentary carbon, nitrogen and phosphorus in Midwestern US reservoirs. *Freshwater Biol.* 59(11):2342–2353.
- Knowlton MF, Jones JR. 2000. Non-algal seston, light, nutrients and chlorophyll in Missouri reservoirs. *Lake Reserv Manage.* 16:322–332.
- Lind OT. 1986. The effect of non-algal turbidity on the relationship of Secchi depth to chlorophyll-a. *Hydrobiologia.* 140:27–35.
- Maindonald J, Braun J. 2007. Data analysis and graphics using R: an example-based approach. 2nd Ed. New York (NY): Cambridge University Press.
- Nürnberg GK. 1996. Trophic state of clear and colored, soft- and hardwater lakes with special consideration of nutrients, anoxia, phytoplankton and fish. *Lake Reserv Manage.* 12:432–447.
- Prepas EE, Planas D, Gibson JJ, Vitt DH, Prowse TD, Dinsmore WP, Halsey LA, McEachern PM, Paquet S, Scrimgeour GJ, et al. 2001. Landscape variables influencing nutrients and phytoplankton communities in Boreal Plain lakes of northern Alberta: a comparison of wetland- and upland-dominated catchments. *Can J Fish Aquat Sci.* 58:1286–1299.
- R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Renwick WH, Vanni MJ, Zhang Q, Patton J. 2008. Water quality trends and changing agricultural practices in a Midwest US watershed, 1994–2006. *J Environ Qual.* 37:1862–1874.
- Richards RP, Baker DB. 2002. Trends in water quality in LEASEQ rivers and streams (northwestern Ohio), 1975–1995. *J Environ Qual.* 31:90–96.
- Schindler DW. 2006. Recent advances in the understanding and management of eutrophication. *Limnol Oceanogr.* 51:356–363.
- Smith VH. 2003. Eutrophication of freshwater and coastal marine ecosystems - a global problem. *Environ. Sci Pollut Res.* 10:126–139.
- Smith VH, Dodds WK, Havens KE, Engstrom DR, Paerl HW, Moss B, Likens GE. 2014. Comment: cultural eutrophication of natural lakes in the United States is real and widespread. *Limnol Oceanogr.* 59:2217–2225.
- Smith VH, Schindler DW. 2009. Eutrophication science: where do we go from here? *Trends Ecol Evol.* 24:201–207.
- Soranno PA, Cheruvilil KS, Webster KE, Bremigan MT, Wagner T, Stow CA. 2010. Using landscape limnology to classify freshwater ecosystems for multi-ecosystem management and conservation. *Bioscience.* 60:440–454.
- Therneau TM, Atkinson B. 2008. rpart: recursive partitioning. R port by B. Ripley.
- Tranvik LJ, Downing JA, Cotner JB, Loiselle SA, Striegl RG, Ballatore TJ, Dillon P, Finlay K, Fortino K, Knoll LB, et al. 2009. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol Oceanogr.* 54:2298–2314.
- Vanni MJ, Andrews JS, Renwick WH, Gonzalez MJ, Noble SJ. 2006. Nutrient and light limitation of reservoir phytoplankton in relation to storm-mediated pulses in stream discharge. *Arch Hydrobiol.* 167:421–445.
- Vollenweider RA, Kerekes J. 1982. Eutrophication of waters. Monitoring, assessment and control. Paris (France): OECD



- Cooperative programme on monitoring of inland waters (eutrophication control); Environment Directorate.
- Wagner T, Soranno PA, Webster KE, Cheruvilil KS. 2011. Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes. *Freshwater Biol.* 56:1811–1824.
- Walker WW. 1984. Trophic state indices in reservoirs. *Lake Reserv Manage.* 1:435–440.
- Whittier TR, Larsen DP, Peterson SA, Kincaid TM. 2002. A comparison of impoundments and natural drainage lakes in the Northeast USA. *Hydrobiologia.* 470:157–171.
- Wickham JD, Stehman SV, Smith JH, Yang L. 2004. Thematic accuracy of the 1992 National Land-Cover Data for the western United States. *Remote Sens Environ.* 91:452–468.

### Supplementary Material

Supplementary Material is available for download via the Inland Waters website, <https://www.fba.org.uk/journals/index.php/IW>:

Supplementary Table S1.