

Quantile regression improves models of lake eutrophication with implications for ecosystem-specific management

YAOYANG XU*, ANDREW W. SCHROTH^{*,†}, PETER D. F. ISLES^{*,‡} AND DONNA M. RIZZO^{*,§}

^{*}Vermont EPSCoR, University of Vermont, Burlington, VT, U.S.A.

[†]Department of Geology, University of Vermont, Burlington, VT, U.S.A.

[‡]Rubenstein School of Environmental and Natural Resources, University of Vermont, Burlington, VT, U.S.A.

[§]School of Engineering, University of Vermont, Burlington, VT, U.S.A.

SUMMARY

1. Although commonly used by those tasked with lake management, the statistical approach of data averaging (DA) followed by ordinary least-squares regression (OLSR) to generate nutrient limitation models is outdated and may impede the understanding and successful management of lake eutrophication.
2. Using a 21-year data set from Lake Champlain as a case study, the traditional DA-OLSR-coupled approach was re-evaluated and improved to quantify the cause–effect relationships between chlorophyll (Chl) and total nitrogen (TN) or total phosphorus (TP).
3. We confirmed that the commonly used DA-OLSR approach results in misleading cause–effect nutrient limitation inferences by illustrating how the process of DA reduces the range of data distribution considered and masks meaningful temporal variation observed within a given period.
4. Our model comparisons demonstrate that using quantile regression (QR) to fit the upper boundary of the response distribution (99th quantile model) is more robust than the OLSR analysis for generating eutrophication models and developing nutrient management targets, as this method reduces the effects of unmeasured factors that plague the OLSR-derived model. Because our approach is statistically in line with the ecological ‘law of the minimum’, it is particularly powerful for inferring resource limitation with broad potential utility to the ecological research community.
5. By integrating percentile selection (PS) with QR-derived model output, we developed a PS-QR-coupled approach to quantify the relative importance of TN and TP reductions in a eutrophic system. Utilising this approach, we determined that the reduction in TP to meet a specific Chl target should be the first priority to mitigate eutrophication in Lake Champlain. The structure of this statistically robust and straightforward approach for developing nutrient reduction targets can be easily adopted as an individual lake-specific tool for the research and management of other lakes and reservoirs with similar water quality data sets.
6. Moreover, the PS-QR-coupled approach developed here is also of theoretical importance to understanding and modelling the interacting effects of multiple limiting factors on ecological processes (e.g. eutrophication) with broad application to aquatic research.

Keywords: data averaging, eutrophication models, least-squares regression, nutrient reduction targets

Introduction

Cultural eutrophication leads to the decline of water quality and degradation of aquatic ecosystems and is a serious environmental problem in many parts of the world

(Smith, Joye & Howarth, 2006). Common symptoms of eutrophic lakes include high turbidity caused by massive algae growth and fish kills caused by severe hypoxia (Schindler, 2006). Harmful algal blooms (HAB) in eutrophic lakes have emerged as an important threat to

Correspondence: Yaoyang Xu, Vermont EPSCoR, University of Vermont, 23 Mansfield Ave, Burlington, VT, 05401, U.S.A.

E-mail: yaoyang.xu@uvm.edu

ecosystem services and human health by fouling drinking water as well as impeding recreational angling, boating and swimming activities (Dodds *et al.*, 2009). Multiple stressors, particularly anthropogenic nutrient loading, contribute to eutrophication, but the relative importance of, and interplay between, various drivers varies among ecosystems (Paerl & Scott, 2010). Variability in the mechanism of eutrophication across diverse lake ecosystems has led to uncertainty in nutrient reduction strategies, and societal struggles with HAB mitigation persist (Smith & Schindler, 2009). For example, management decisions related to eutrophication have commonly been made from a regulatory guidance perspective (e.g. nutrient criteria development; U.S. EPA, 2000) and models generated from multiple-lake data sets (e.g. a chlorophyll-nutrient model; Phillips *et al.*, 2008), neither of which is necessarily appropriate for individual, lake-specific eutrophication management (Mooij *et al.*, 2010; Huo *et al.*, 2013). Therefore, eutrophication management efforts should be directed towards ecosystem-specific strategies when the available data provide sufficient power to develop and validate individual lake-eutrophication models.

While the reduction in anthropogenic nutrient inputs has been widely recognised as the primary tool for mitigating eutrophication, an important debate in policy forums concerns the optimal design of nitrogen and phosphorus reductions (Conley *et al.*, 2009a; Schelske, 2009). Due to the high cost of controlling nitrogen inputs and the potential for nitrogen fixation by some cyanobacteria, heavy emphasis has been placed on decreasing phosphorus inputs to restore eutrophic ecosystems (Conley *et al.*, 2009b; Schindler & Hecky, 2009). However, dual nutrient reduction strategies for lake management have also been advocated because phosphorus-only reduction strategies may be insufficient to control eutrophication, and increased nitrogen inputs could exacerbate eutrophication problems (Bryhn & Håkanson, 2009; Jacoby & Frazer, 2009). In fact, neither phosphorus-only reduction nor dual nutrient reduction may be universally effective, and it is important to determine which of these strategies is useful in the context of a specific lake. Prior to this policy forum, the U.S. EPA (2000) proposed two percentile-based options to help develop regional nutrient reduction targets. The 75th percentile option defines reference conditions based on data from lakes with most 'natural' condition, whereas the 25th percentile option defines reference targets based on sampling all lakes within the region, including those with considerably less than 'natural' condition (U.S. EPA, 2000). However, the selection of these two percentile-based options is, to some extent, qualitative because determin-

ing lake condition as 'natural and minimally impacted' is unavoidably subjective (U.S. EPA, 2000) and because the degree of impact in 'unnatural' lakes may vary widely. With the growth of lake long-term monitoring data, there is a golden opportunity to quantify nutrient reduction targets by integrating the percentile selections with ecologically relevant lake-specific nutrient models.

The application of empirical eutrophication models relating chlorophyll (Chl) to total nitrogen (TN) or total phosphorus (TP) has been a common approach for determining nutrient limitation and developing strategies to control anthropogenic eutrophication since the 1960s (e.g. Sakamoto, 1966; Dillon & Rigler, 1974; Stauffer, 1991). However, several methodological issues associated with these common models contribute to bias and uncertainty in the inference of nutrient limitation and the determination of nutrient targets. We will demonstrate that these can be avoided easily with the adaptation of recent advances in statistical limnology. While Chl-nutrient modelling derived from multi-lake data sets is necessary and practical when water quality data from individual lakes are not available (e.g. Malve & Qian, 2006; Higgins *et al.*, 2011; Cha, Stow & Bernhardt, 2013), the potential for inaccurate predictions using multiple-lake regression models on individual lakes has been recognised since at least the 1980s (Smith & Shapiro, 1981). Furthermore, the proliferation of large monitoring data sets in the past three to four decades makes it possible in many cases to develop a lake-specific approach for model development (Cardoso *et al.*, 2007). In addition, averaging individual observations spatially (within a lake or sub-basin) or temporally (across seasons or monitoring years) is commonly applied in the generation of eutrophication models (e.g. Phillips *et al.*, 2008; Wagner *et al.*, 2011); although the negative effects of data aggregation on the predictive ability were documented in the 1990s (Jones, Knowlton & Kaiser, 1998) and the mathematical issue of data averaging (DA) in the log transformation was addressed in the 2000s (Stow, Reckhow & Qian, 2006). Furthermore, ordinary least-squares regression (OLSR) remains a common statistical method for producing eutrophication models (e.g. Håkanson *et al.*, 2005; Huszar *et al.*, 2006; Wang *et al.*, 2014), despite wide acknowledgement of the fundamental conflict between the underlying statistical assumptions and the ecological concept of limiting factors in limnology (Kaiser, Speckman & Jones, 1994). Given these methodological problems, an important next step for modellers and management communities is to jointly recognise the need to advance statistical modelling of lake-specific eutrophication.

Quantile regression (QR), originally derived from econometric science (Koenker & Bassett Jr, 1978), was

introduced in statistical ecology (Cade, Terrell & Schroeder, 1999) as an alternative to OLSR to quantify better cause–effect relationships between ecological variables across the range of the response variable distribution. In contrast to the OLSR-derived model, which fails to fit a predictive relationship between the response variable and the measured factor given the inclusion of other unmeasured effects, a QR-derived 99th quantile model, which enables best fit of the upper boundary of the response variable distribution, provides a more robust prediction of the response maxima at a given value of the measured variable by controlling for the potential effects of other unmeasured factors (Cade & Noon, 2003). Fundamentally, using the 99th quantile model to fit the upper boundary follows the law of the minimum in ecology, which states that ecological processes are controlled by the scarcest resource, not the total amount of resources available (Cade *et al.*, 1999; Cade & Guo, 2000). Quantile regression has recently been expanded into many ecological applications (e.g. Sterck *et al.*, 2014; Wagner, Harmon & Seehausen, 2014), but has not yet been widely adopted by statistical limnologists or water quality managers (Abell *et al.*, 2012; Carvalho *et al.*, 2013), despite the noted theoretical shortcomings of the OLSR-derived eutrophication models (Kaiser *et al.*, 1994).

Using long-term monitoring data from Lake Champlain (1992–2012), this study shows the ecological and practical advantages of adopting the QR analysis to systematically model lake-specific eutrophication. The following steps were taken. First, eutrophication models were developed using the DA-OLSR modelling approach to infer nutrient limitation and develop management targets for Lake Champlain. Next, the advantages of the QR analysis, relative to OLSR for modelling nutrient limitation and developing management targets, are highlighted using all Lake Champlain individual observations. Lastly, we coupled the percentile selection (PS) with the QR analysis to develop a robust approach that quantifies the relative importance of nitrogen and phosphorus reductions. The structure of the PS-QR-coupled approach and associated nutrient criteria development (Fig. 1a–d) is broadly applicable to other systems where similar water quality data sets exist.

Methods

Data set source

Lake Champlain is a large lake located in north-eastern America in the continental rift valley between the Green (Vermont) and Adirondack (New York) mountains

(O'Donnell *et al.*, 2013). The lake is 122 m deep, 19 km in width at its maximum and 194 km in length, with a volume of 26 km³ and an area of 1127 km² (Smeltzer, Shambaugh & Stangel, 2012). The spatially diverse habitat is characterised by complex bathymetry with over 70 islands, a shoreline of over 800 km and many semi-isolated bays (Chen *et al.*, 2012). The drainage basin has mean annual rainfall and snowfall of 84.5 and 155 cm, and average annual minimum and maximum temperatures of –13.9 and 26.0 °C in January and July, respectively (Levine *et al.*, 2012). Compared to the lake surface, the catchment is large (a 19 : 1 ratio) and includes 11 major rivers and many smaller tributaries (Marsden & Langdon, 2012).

The data for this case study were collected under the Long-Term Water Quality and Biological Monitoring Program on Lake Champlain (LTMP) by the Vermont Department of Environmental Conservation and the New York Department of Environmental Conservation since 1992 (Vermont DEC & York DEC, 2013). Each year, 15 sampling locations spanning the lake were monitored approximately fortnightly from May to early November. Details of sampling methods and quality controls are documented in the Quality Assurance Project Plan (VTDEC (Vermont Department of Environmental Conservation) & NYDEC (New York State Department of Environmental Conservation), 2013), and water quality monitoring data are freely available for scientific research at the web site: http://www.vtwaterquality.org/lakes/htm/lp_longterm.htm. For this study, we downloaded 3545 paired chlorophyll (Chl, mg m⁻³), total nitrogen (TN, mg m⁻³) and total phosphorus (TP, mg m⁻³) observations for each of the 15 stations from 1992 to 2012.

Model procedures

Following the conventional DA-OLSR-coupled approach for nutrient limitation modelling, the 21-year Chl and nutrient values from the entire data set ($n = 3545$) were first averaged annually for each of the 15 sampling stations, log₁₀-transformed and analysed using OLSR to generate the DA-OLSR models of annual means. The subset of 2130 observations that make up our 'summer' data set use the definition of a summer season in the Lake Champlain region as the period from June to August in regional climate studies (Guilbert *et al.*, 2014). Similar to the analysis of the entire data set, the summer observations were also analysed using the DA-OLSR-coupled approach to produce the DA-OLSR summer mean models.

Alternatively, all 3545 individual Chl and nutrient observations were log₁₀-transformed and analysed using

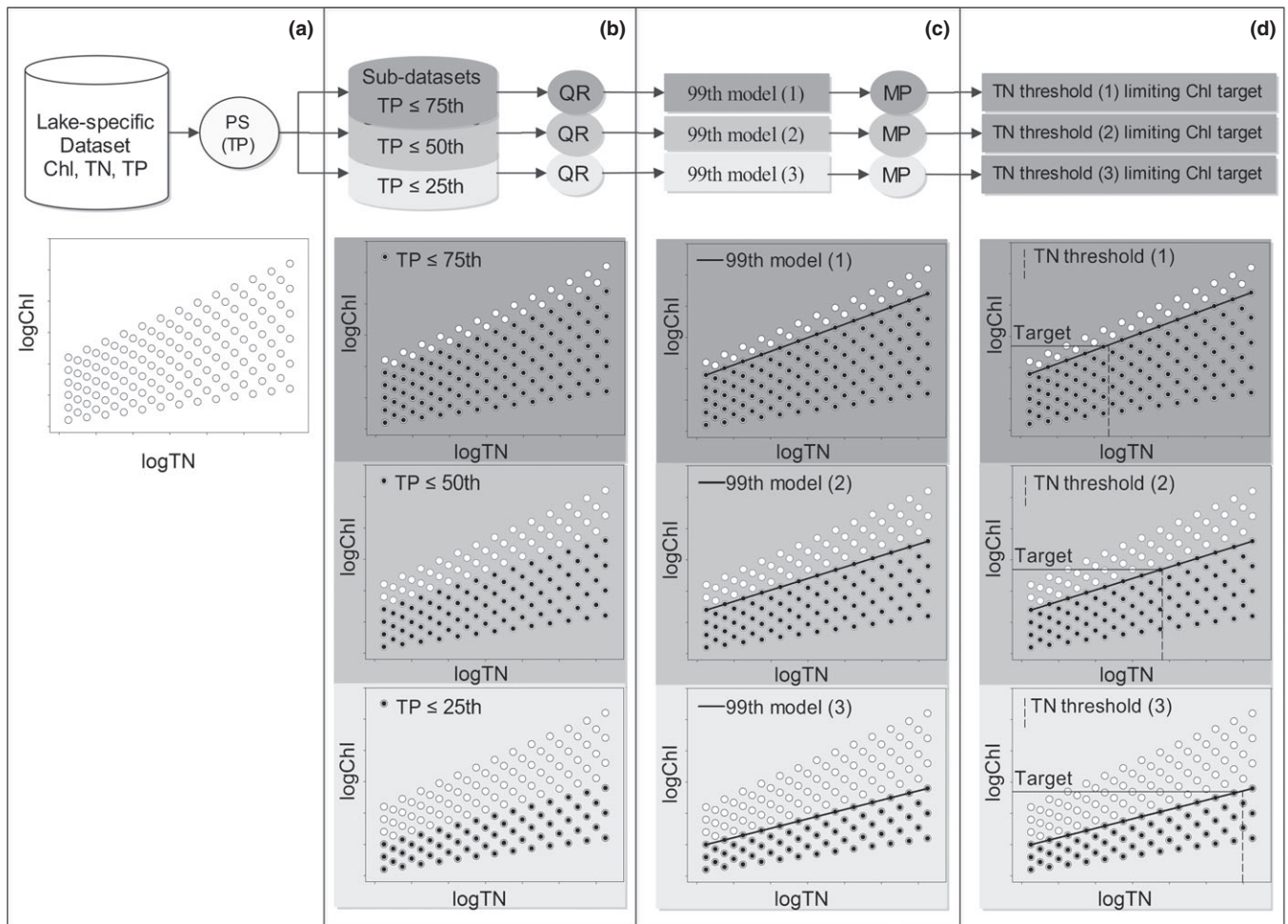


Fig. 1 The coupled framework for modelling lake-specific eutrophication illustrating the percentile selection (PS, a) of nutrient (e.g. total phosphorus, TP), quantile regression of subdata sets (e.g. chlorophyll-total nitrogen; logChl-logTN, b) and the 99th models prediction (MP, c) for TN thresholds (d) that limit a specific Chl target.

OLSR and QR to generate the OLSR models and the quantile models. Following the QR methodology of Koenker & Machado (1999), coefficients of determination (R^1) were computed as follows: $R^1 = 1 - (SUM_{(F)}:SUM_{(R)})$, where $SUM_{(F)}$ is the sum of the weighted absolute deviations minimised in estimating each of the *full* parameter models (i.e. $\log\text{Chl} = b_0 + b_1\log\text{TN} + \varepsilon$ and $\log\text{Chl} = b_0 + b_1\log\text{TP} + \varepsilon$), and $SUM_{(R)}$ is the sum of the weighted absolute deviations minimised in estimating each of the corresponding reduced parameter models ($\log\text{Chl} = b_0 + \varepsilon$). The QR-derived R^1 values were further converted to the unit of measure (R^2) comparable to the OLSR-derived R^2 values following the formula (McKean & Sievers, 1987; Schooley & Wiens, 2005): $R^2 = 1 - (1 - R^1)^2$. Moreover, the Akaike information criterion difference was computed according to Cade, Noon & Flather (2005) as follows: $\Delta\text{AIC} = \text{AIC}_{(R)} - \text{AIC}_{(F)}$, where $\text{AIC}_{(F)}$ and $\text{AIC}_{(R)}$ are Akaike information criteria for the *full* parameter models

(i.e. $\log\text{Chl} = b_0 + b_1\log\text{TN} + \varepsilon$ and $\log\text{Chl} = b_0 + b_1\log\text{TP} + \varepsilon$) and the corresponding reduced parameter model ($\log\text{Chl} = b_0 + \varepsilon$), respectively. More details of the QR methodology targeted to ecologists can be found in Cade & Guo (2000), Cade & Noon (2003), and Cade *et al.* (2005). The OLSR analysis was performed with the SPSS 13.0 software, while the QR analysis used the R (R Development Core Team, 2014) package quantreg version 5.05 (Koenker *et al.*, 2013).

To quantify the effect of TP percentile selection (PS) on the limitation model and management target of nitrogen (i.e. 99th quantile model $\log\text{Chl} = b_0 + b_1\log\text{TN} + \varepsilon$), we developed a coupled PS-QR approach as follows (Fig. 1): (i) the 75th, 50th and 25th percentiles of TP concentrations for the entire water quality data set (Fig. 1a) were selected to define three subdata sets, respectively (Fig. 1b); (ii) applying the QR analysis to each of these three subdata sets (Fig. 1b), the upper boundaries of

Chl-TN distributions at a log-log scale are fit by the 99th quantile model $\log\text{Chl} = b_0 + b_1\log\text{TN} + \varepsilon$ (Fig. 1c); and (iii) the 99th quantile models for three subdata sets (Fig. 1c) predict and compare the TN concentration threshold targets (Fig. 1d) for controlling a specific Chl target under decreasing percentiles of TP concentration. Similarly, we repeated the PS-QR-coupled approach to quantify the effect of TN percentiles selection on the limitation model and management target of phosphorus (i.e. 99th quantile model $\log\text{Chl} = b_0 + b_1\log\text{TP} + \varepsilon$).

Results

The DA-OLSR-derived models

The 3545 observations comprising the entire data set represent a wide range of chlorophyll and nutrient concentrations within Lake Champlain (Table 1). Based on the 1992–2012 annual means (Chl_{am} , TN_{am} and TP_{am} , diamonds of Fig. 2a and b) for each of the 15 sampling

stations at the \log_{10} scale, the DA-OLSR-coupled analysis produced the annual models: $\log\text{Chl}_{\text{am}} = 1.47(\pm 0.12)\log\text{TN}_{\text{am}} - 3.14(\pm 0.31)$ ($n = 284$, $R^2 = 0.351$, $\text{AIC} = -138$, $P < 0.01$, Fig. 2a) and $\log\text{Chl}_{\text{am}} = 0.74(\pm 0.04)\log\text{TP}_{\text{am}} - 0.26(\pm 0.05)$ ($n = 284$, $R^2 = 0.611$, $\text{AIC} = -284$, $P < 0.01$, Fig. 2b). Moreover, the summer ($n = 2130$) data set shows a wide range of chlorophyll and nutrient concentrations (Table 1). The DA-OLSR-coupled analysis of the summer data set (2130 observations, open circles of Fig. 3a and b) generates the empirical models of summer means (Chl_{sm} , TN_{sm} and TP_{sm} , diamonds of Fig. 3a and b): $\log\text{Chl}_{\text{sm}} = 1.33(\pm 0.12)\log\text{TN}_{\text{sm}} - 2.78(\pm 0.32)$ ($n = 284$, $R^2 = 0.287$, $\text{AIC} = -92$, $P < 0.01$, Fig. 3a) and $\log\text{Chl}_{\text{sm}} = 0.72(\pm 0.04)\log\text{TP}_{\text{sm}} - 0.24(\pm 0.05)$ ($n = 284$, $R^2 = 0.545$, $\text{AIC} = -219$, $P < 0.01$, Fig. 3b).

Ordinary and quantile regressions

The OLSR analysis of the paired \log_{10} -transformed data for 3545 individual observations produces the OLSR

Table 1 Statistical summary of chlorophyll (Chl, mg m^{-3}), total nitrogen (TN, mg m^{-3}) and total phosphorus (TP mg m^{-3}) for the entire $n = 3545$ data set and the summer $n = 2130$ data set

Data set	Variable	N	Mean	SD	Max	75th	50th	25th	Min
Entire	Chl	3545	6.1	7.0	116.4	7.1	4.2	2.8	0.4
	TN	3545	417.8	143.9	1720.0	450.0	390.0	340.0	110.0
	TP	3545	24.2	17.7	235.0	32.0	16.4	12.0	5.0
Summer	Chl	2130	6.2	7.6	116.4	6.8	4.1	2.7	0.5
	TN	2130	416.2	144.9	1710.0	440.0	380.0	340.0	110.0
	TP	2130	24.4	17.9	235.0	32.0	16.8	12.3	5.0

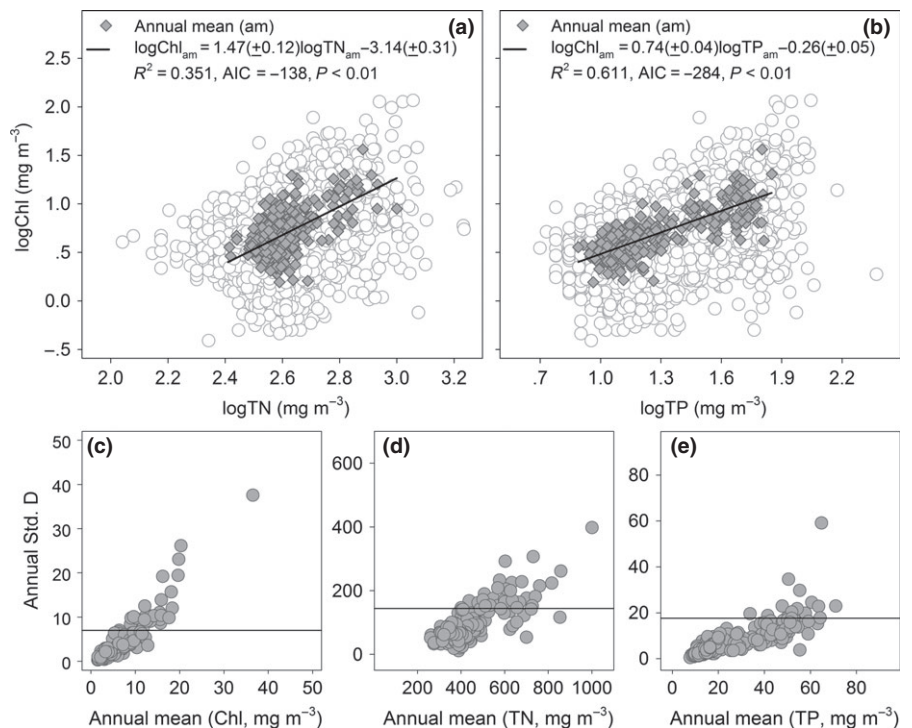


Fig. 2 Scatter plots of chlorophyll (Chl) against total nitrogen (TN, a) and total phosphorus (TP, b) for all $n = 3545$ observations of Lake Champlain with the DA-OLSR-derived models of annual means (diamonds), as well as the standard deviation (SD) against annual means of Chl (c), TN (d) and TP (e). The horizontal line (c–e) indicates the standard deviations of all $n = 3545$ observations.

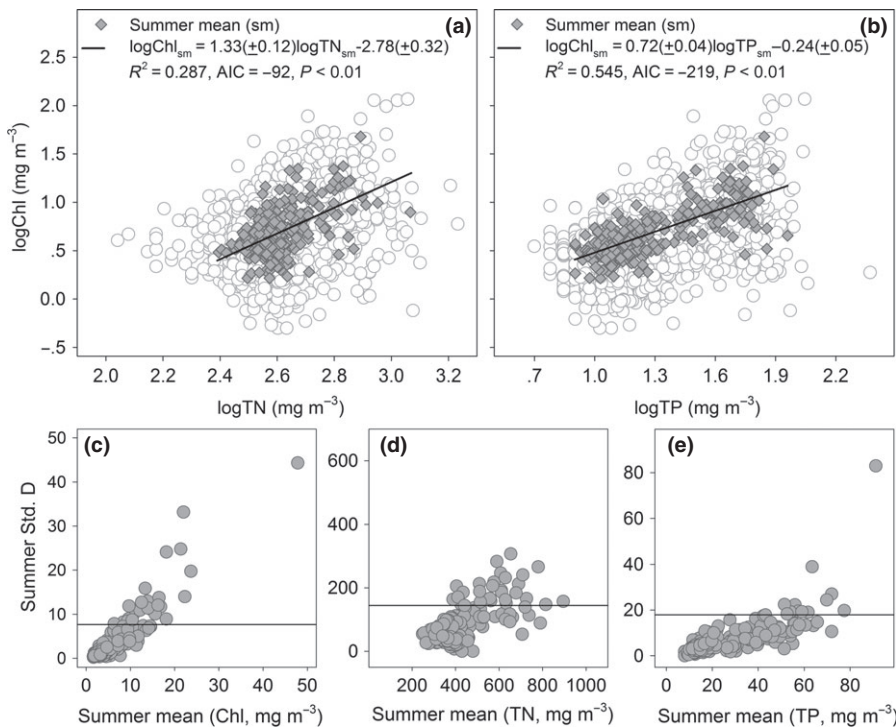


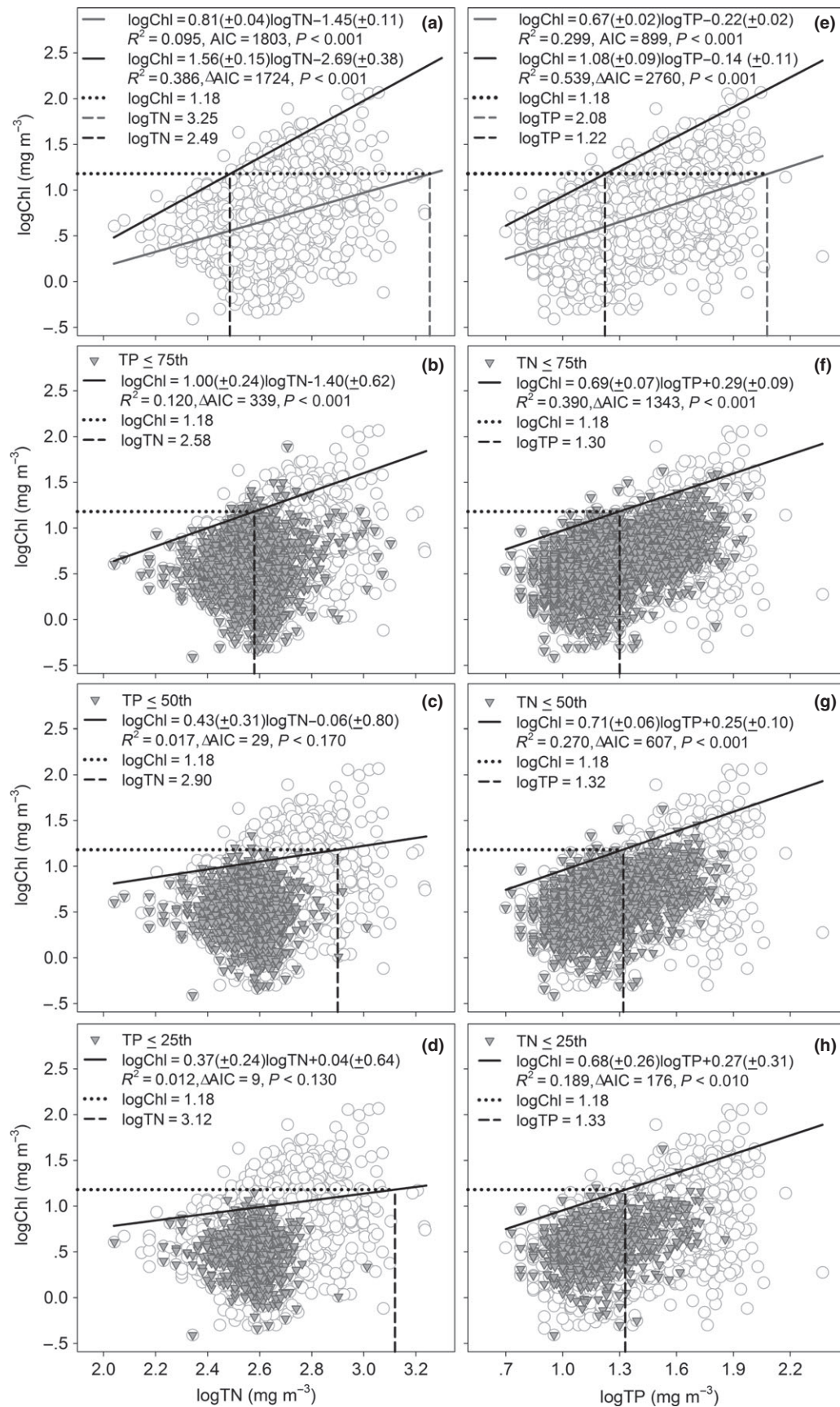
Fig. 3 Scatter plots of chlorophyll (Chl) against total nitrogen (TN, a) and total phosphorus (TP, b) for summer $n = 2130$ observations of Lake Champlain with the DA-OSLR-derived models of summer means (diamonds), as well as the standard deviation (SD) against summer means of Chl (c), TN (d) and TP (e). The horizontal line (c–e) indicates the standard deviations of summer $n = 2130$ observations.

models: $\log\text{Chl} = 0.81(\pm 0.04)\log\text{TN} - 1.45(\pm 0.11)$ ($n = 3545$, $R^2 = 0.095$, $\text{AIC} = 1803$, $P < 0.001$; dark grey line of Fig. 4a) and $\log\text{Chl} = 0.67(\pm 0.02)\log\text{TP} - 0.22(\pm 0.02)$ ($n = 3545$, $R^2 = 0.299$, $\text{AIC} = 899$, $P < 0.001$; dark grey line of Fig. 4e). Using these paired data, the QR analysis results in increased coefficients of determination (R^2 , Fig. 5a) from 0.001 to 0.386, ΔAIC value from 4 to 1724 (Fig. 5b), significance level from $P < 1.000$ to $P < 0.001$ (Fig. 5c) and slopes (b_1 , Fig. 5d) from $-0.11(\pm 0.11)$ to $1.56(\pm 0.15)$ as well as decreased intercepts (b_0 , Fig. 5e) from $0.35(\pm 0.28)$ to $-2.69(\pm 0.38)$ with increasing quantiles for the $\log\text{Chl} = b_0 + b_1\log\text{TN} + \epsilon$ model. The coefficients of determination, ΔAIC value, significant level, slopes and intercepts for the $\log\text{Chl} = b_0 + b_1\log\text{TP} + \epsilon$ model increase gradually from 0.022 to 0.539 (R^2 , Fig. 5f), 76 to 2760 (Fig. 5g), $P < 0.064$ to $P < 0.001$ (Fig. 5h), $0.21(\pm 0.11)$ to $1.08(\pm 0.09)$ (b_1 , Fig. 5i) and $-0.33(\pm 0.14)$ to $-0.14(\pm 0.11)$ (b_0 , Fig. 5j) with increasing quantiles. The upper Chl–TN and Chl–TP boundaries are best fit by the 99th quantile models $\log\text{Chl} = 1.56(\pm 0.15)\log\text{TN} - 2.69(\pm 0.38)$ ($R^2 = 0.386$, $\Delta\text{AIC} = 1724$, $P < 0.001$, dark line of Fig. 4a) and $\log\text{Chl} = 1.08(\pm 0.09)\log\text{TP} - 0.14(\pm 0.11)$ ($R^2 = 0.539$, $\Delta\text{AIC} = 2760$, $P < 0.001$, dark line of Fig. 4e), respectively.

Percentile selection and model estimations

A statistical summary (Table 1) shows the 75th, 50th and 25th percentiles of the 3545 observations with TP concentrations of 33.1, 17.0 and 12.3 mg m^{-3} and TN concentrations of 450.0, 390.0 and 340.0 mg m^{-3} . The PS-QR-coupled analysis further estimated the 99th quantile model parameters for both the $\log\text{Chl} = b_0 + b_1\log\text{TN} + \epsilon$ under reduced TP concentrations (triangles of Fig. 4b–d) and the $\log\text{Chl} = b_0 + b_1\log\text{TP} + \epsilon$ under reduced TN concentrations (triangles of Fig. 4f–h). The 99th quantile model $\log\text{Chl} = b_0 + b_1\log\text{TN} + \epsilon$ exhibits a dramatic decline in coefficients of determination (R^2) from 0.386 to 0.012, ΔAIC value from 1724 to 9, significant level from $P < 0.001$ to $P < 0.130$ and slopes from $1.56(\pm 0.15)$ to $0.37(\pm 0.24)$ as well as a dramatic increase in the intercept from $-2.69(\pm 0.38)$ to $0.04(\pm 0.64)$ with reducing TP concentration (dark line, Fig. 4a–d). For the 99th quantile model $\log\text{Chl} = b_0 + b_1\log\text{TP} + \epsilon$, the coefficients of determination, ΔAIC value and significant level decrease from 0.539 to 0.189, 2760 to 176 and $P < 0.001$ to $P < 0.010$ overall with reducing TN concentrations, respectively. When the TN concentration is

Fig. 4 The OSLR-derived models (dark grey line) and the 99th quantile models (dark line) of Chl–TN distribution at log10 scale for all 3545 observations (a) and these observations under decreasing TP concentration (b: 75th, c: 50th, d: 25th; triangles), as well as Chl–TP distribution at log10 scale for all 3545 observations (e) and these observations under decreasing TN concentration (f: 75th, g: 50th, h: 25th; triangles). The thresholds of TN and TP for limiting a specific Chl target (e.g. 15.0 mg m^{-3} , $\log\text{Chl} = 1.18$; dotted line) are indicated by the dark grey-dashed line (the OSLR-derived model prediction, a and e) and the dark-dashed line (the 99th quantile model prediction, a–h).



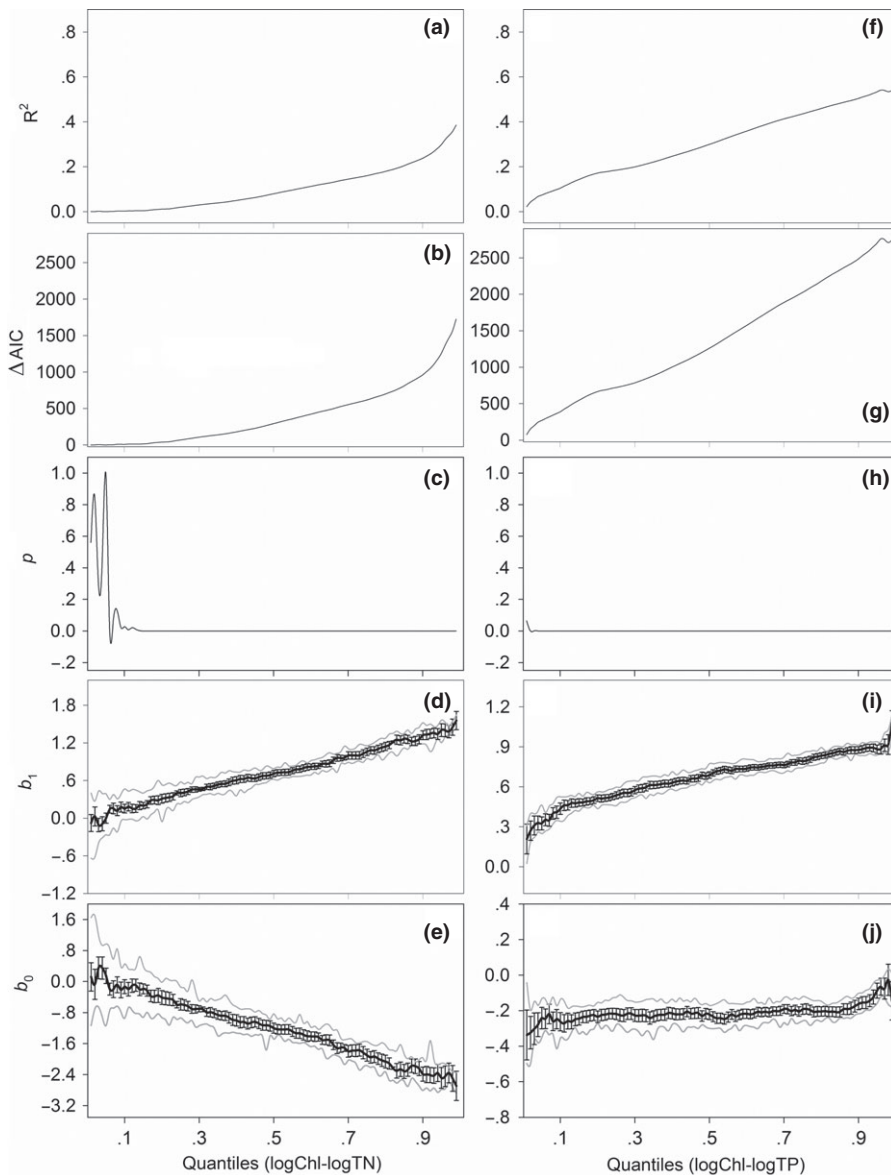


Fig. 5 The coefficient of determination (R^2), Akaike information criterion (ΔAIC), significance level ($P < .$), slopes (b_1 , grey lines: 95% confidence interval; black bar: standard error) and intercepts (b_0 , grey lines: 95% confidence interval; black bar: standard error) for $\log\text{Chl} = b_0 + b_1\log\text{TN} + \varepsilon$ (a–e) and $\log\text{Chl} = b_0 + b_1\log\text{TP} + \varepsilon$ (f–j) with increasing quantiles of all 3545 observations.

reduced from 100th to 75th percentile, the slope decreases from 1.08 to 0.69 and the intercept increases from -0.14 to 0.29 (dark line, Fig. 4e,f). However, when TN is further reduced from 75th to 25th percentile, the slope and intercept remain nearly unchanged (dark line, Fig. 4f–h).

Discussion

Issues of the conventional approach

Using the conventional DA-OLSR-coupled approach, empirical linear models describing the relationship between Chl and TN and TP were generated for the entire ($n = 3545$) data set. In general, the coefficients of determination were higher for the DA-OLSR-derived

model of Chl and TP than for that of Chl and TN. These results appear consistent with the canonical inference of statistical limnology, which used the relatively high coefficient of determination to indicate phosphorus limitation (e.g. Phillips *et al.*, 2008; Wang *et al.*, 2008). Although it may be unreasonable to expect either of these DA-OLSR models based on annual means to provide a good fit for the entire $n = 3545$ data set, the variation being masked by the process of DA is likely to be critical for robust inferences of nutrient limitation that are in line with the ecological law of the minimum. On the one hand, the DA dramatically reduced the range of distribution for the entire Lake Champlain data set, while on the other hand, the process of DA also aggregated the multiple annual observations for each year at each of the 15 sampling locations into the simple surrogate for Chl, TN and

TP. Indeed, such an approach masked the temporal variation of these predictor and response variables within the annual time period. Furthermore, some of this masked temporal variation was higher than the total variation of the entire Lake Champlain data set, as shown by the scatter plots of standard deviation versus mean. Furthermore, these same problematic issues identified in our $n = 3545$ data set analysis are revealed when multiple summer observations are averaged prior to the OLSR analysis using the DA-OLSR-coupled approach on the $n = 2130$ summer data set. Since DA reduces the range of the data distribution under consideration and masks the temporal variation observed within the annual or summer period, it is reasonable to believe that the DA-OLSR-coupled approach is not applicable for statistical inference of phosphorus limitation in Lake Champlain and that eutrophication models derived from such an approach are fundamentally flawed.

The high level of uncertainty masked by the DA is likely to be present for many published models that average data prior to OLSR analysis (e.g. Higgins *et al.*, 2011; Wagner *et al.*, 2011). As a result, those published models may have been applied inappropriately when statistically inferring nutrient limitation for their lake ecosystems, given the strong influence of DA across multi-observations (Mooij *et al.*, 2010; Stow & Cha, 2013). Furthermore, an inaccurate inference of nutrient limitation may result in similarly flawed nutrient management targets for the modelled lake ecosystems. In addition to these statistical inference issues, the direct comparison of model parameters among different studies is not possible due to the DA uncertainty effects on model estimations, potentially contributing to the ongoing debate on eutrophication management strategy. While the violation that DA imposes on the statistical assumptions has been theoretically addressed (Jones *et al.*, 1998; Stow *et al.*, 2006), the incorporation of straightforward advances in statistical science to avoid this issue has not been practically or broadly applied to the common eutrophication modelling approach within the international limnological community. The obvious potential for misleading inferences of nutrient limitation and associated targets generated by the conventional DA-OLSR-derived Chl-nutrient models serves as strong motivation to derive a more robust statistical modelling approach that could be broadly utilised for setting lake-specific nutrient criteria.

Revised statistical approach

In the light of the noted flaws associated with averaging annual or summer observations of Chl, TN and TP, we

generated the OLSR-derived models using all 3545 observations. Given the common use of OLSR-derived R^2 values as indicative of a cause-effect relationship, or lack thereof (e.g. Prairie, Duarte & Kalff, 1989; Stow & Cha, 2013), neither TN nor TP would be considered a limiting nutrient of Chl within Lake Champlain based on OLSR-derived model output. However, this model result does not appear consistent with the general consensus that reducing nutrient inputs is the most effective way to mitigate cultural eutrophication, as it is well known that long-term nutrient loading from the catchment has caused eutrophication in Lake Champlain (Levine *et al.*, 2012; Smeltzer *et al.*, 2012). In fact, one of the limitations of using the OLSR-derived R^2 values to infer nutrient limitation is that the OLSR analysis cannot eliminate the effects of drivers of Chl concentration other than TN or TP. This limitation is clearly evident in the Lake Champlain analysis, as the distribution of Chl-TN (or Chl-TP) suggests that Chl responds differently to a given TN (or TP) concentration. To be consistent with the ecological 'law of the minimum', nitrogen can be regarded as the only limiting factor when Chl is equal to or near its maximum value at a given TN concentration, while a factor other than nitrogen may be the active limiting constraint when Chl is lower than its maximum value. Similarly, phosphorus can be identified as the only limiting factor when Chl response to a given TP concentration is at or near its maximum value, while a factor other than phosphorus may be the active limiting constraint when this response falls below its maximum value. The OLSR analysis, however, focuses on the mean of other unmeasured factor-limited Chl responses as a function of a given TN (or TP) concentration and generates empirical linear relationships that are not useful for inferring nitrogen and phosphorus limitations in Lake Champlain. This suggests that a better statistical approach to reduce the potential impact of unmeasured factors on cause-effect models of Chl nutrient is clearly warranted in the case of Lake Champlain.

Furthermore, it is logical that the impact of unmeasured factors on the OLSR-derived Chl-nutrient models will cascade to the generation of flawed nutrient criteria. Indeed, when the reduced Chl target concentration is set at 15 mg m^{-3} (USEPA, 2000; $\log \text{Chl} = 1.18$, dotted line of Fig. 4a and e), the OLSR-derived models would predict the threshold of TN and TP concentrations to be 1793.6 mg m^{-3} ($\log \text{TN} = 3.25$, dark grey-dashed line of Fig. 4a) and 119.8 mg m^{-3} ($\log \text{TP} = 2.08$, dark grey-dashed line of Fig. 4e), respectively. Hypothetically, Lake Champlain managers could then implement strategies to reach those target thresholds of TN

(1793.6 mg m⁻³) and TP (119.8 mg m⁻³). However, a more comprehensive analysis indicates that reducing nutrients to those targets would not produce the desired suppression of Chl concentration. The spurious predictions of the OLSR-derived models are more obviously revealed by the Chl observations, which distribute above its reduced target line (logChl = 1.18, dotted line of Fig. 4a and e), despite having TN and TP concentrations less than the predicated threshold values (logTN = 3.25, dark grey-dashed line of Fig. 4a; logTP = 2.08, dark grey-dashed line of Fig. 4e). As such, the inaccurate predictions we illustrated here confirm the problematic application of the OLSR analysis in the development of eutrophication management targets.

Given that other unmeasured factors (e.g. temperature) cannot be managed and that eutrophication management considers nutrients (TN and TP) as the controlling factors of Chl, it is intuitive that managers adopt a statistical approach capable of inferring nutrient limitation. Considering the fact that the OLSR-derived R^2 values measure goodness of fit over the entire Chl-TN (or Chl-TP) distribution of data yet involve the effect of other unmeasured factors, we optimised the inference of nutrient limitations using the QR-derived R^2 values, which measure goodness of fit for a particular quantile of Chl-TN (or Chl-TP) distribution. The increased QR-derived R^2 values reveal that other unmeasured factors constraining Chl below the upper boundary of the Chl-TN (or Chl-TP) distribution can be excluded gradually with increasing quantiles. The highest QR-derived R^2 value at the 99th quantile shows that the 99th quantile model represents a best fit for the upper boundary of the Chl-TN (or Chl-TP) distribution. These results are completely consistent with the assumptions that nitrogen (or phosphorus) limitation within the system is captured at the upper boundary of the Chl-TN (or Chl-TP) distribution (e.g. Jones, Obrecht & Thorpe, 2011). Moreover, the higher R^2 values for the 99th quantile models, relative to the OLSR-derived models, confirm that the QR analysis is more robust than OLSR for inferring nutrient limitations.

Equipped with a more robust statistical inference of nutrient limitation, another advantage of using the QR approach is that the 99th quantile models can provide more reliable prediction than the conventional OLSR-derived models of TN and TP thresholds, needed to control the Chl target within Lake Champlain. To constrain Chl below a specific target (e.g. 15 mg m⁻³), predicted TN and TP thresholds using the 99th quantile models are, respectively, 305.7 mg m⁻³ (logTN = 2.49, dark-dashed line of Fig. 4a) and 16.7 mg m⁻³ (logTP = 1.22, dark-dashed of Fig. 4e). These predictions are consistent

with the Chl-TN and Chl-TP distributions, in which Chl concentrations are always lower than the target when the concentrations of TN and TP are less than the predicted thresholds. The accurate TN and TP threshold predictions illustrated here confirm the advantage of the QR analysis in the development of nutrient management targets. Moreover, the QR analysis could be applicable to generate other quantile models (e.g. 95th and 90th) as it would be acceptable for lake managers, given that the 99th quantile model is regarded as too stringent for practical lake management. Thus, our analyses provide practical insight into how eutrophication models commonly used by research communities and management entities can evolve to incorporate straightforward statistical advances in ecology, therefore improving their utility to infer nutrient limitations and determine effective nutrient thresholds.

Integration of percentile selection and quantile regression

After clear demonstration of the QR analysis advantages relative to OLSR for Chl-nutrient model and threshold development, the logical next step is to build a robust approach to detect the relative importance of nitrogen and phosphorus in controlling Chl concentration for individual lake ecosystems. Using the case of Lake Champlain, we illustrate the power of using a PS-QR-coupled approach for developing lake-specific nutrient management strategies. When applying the PS-QR-coupled approach to the Lake Champlain data set, the reduced R^2 values detected for the 99th quantile models $\log\text{Chl} = b_0 + b_1\ln\text{TN} + \varepsilon$ (dark line of Fig. 4a–d) suggest that nitrogen impact on Chl decreases when TP concentrations are held within the 25th percentile. Predictions from these 99th quantile models reveal that the thresholds of TN concentration for limiting a specific Chl target increase dramatically when reducing TP concentration. In other words, the need to reduce nitrogen inputs could become less if Lake Champlain managers were able to effectively reduce phosphorus inputs. For example, the concentration of TN could be relaxed to 1312.6 mg m⁻³ (logTN = 3.12, dark-dashed line of Fig. 4d), as long as TP concentration in Lake Champlain can be reduced to the 25th percentile (i.e. 12.0 mg m⁻³). As such, the PS-QR-coupled approach, which involves in the selection of TP percentiles and the estimation of nitrogen models, could provide robust support for developing a TP reduction-priority strategy for mitigating Lake Champlain eutrophication.

Similarly, we applied the PS-QR-coupled approach to infer phosphorus limitation and predict management tar-

gets under TN reduction. In the present case, the extent to which the R^2 value decreases for the 99th quantile model $\log\text{Chl} = b_0 + b_1\ln\text{TP} + \varepsilon$ with reducing TN concentrations is less than that for the 99th quantile model $\log\text{Chl} = b_0 + b_1\ln\text{TN} + \varepsilon$ with reducing TP concentrations. Unlike TP reduction, which dramatically affects nitrogen limitation, TN reduction exerts a less marked effect on phosphorus limitation. For the 99th quantile model $\log\text{Chl} = b_0 + b_1\ln\text{TP} + \varepsilon$, the decreased slope suggests that TN reductions from the 100th to 75th percentile increase the effectiveness of moderate TP reductions, but the consistent slope indicates that TN reductions below the 75th percentile are unlikely to be effective. By applying these 99th quantile models $\log\text{Chl} = b_0 + b_1\ln\text{TP} + \varepsilon$ to predict the threshold of TP for a specific Chl target, we found that TP concentration can be relaxed slightly from 16.7 mg m^{-3} ($\log\text{TP} = 1.22$, dark-dashed line of Fig. 4e) to 19.8 mg m^{-3} ($\log\text{TP} = 1.30$, dark-dashed line of Fig. 4f) when reducing TN concentration into 75th percentiles and that it must be controlled to $<21.4 \text{ mg m}^{-3}$ ($\log\text{TP} = 1.33$, dark-dashed line of Fig. 4h) even if TN concentration is reduced into the 25th percentile. These findings further suggest that reducing TP concentration remains the most important strategy for achieving Chl targets, although the effectiveness of these reductions increases somewhat with moderate decreases in TN. Furthermore, this analysis clearly illustrates how the PS-QR-coupled approach might be utilised by researchers to develop multiple strategies for reaching Chl targets in eutrophic systems, potentially giving management and stakeholder communities increased flexibility as they try to reach their water quality goals.

From applying the PS-QR-coupled approach to Lake Champlain historical data, we conclude that TP reduction should be the management priority, although dual nutrient reductions would be better for eutrophication mitigation of this lake and allow for flexible TP targets. However, we must emphasise that this is an ecosystem-specific conclusion, derived from an ecosystem-specific data set and empirical model. Since it is unrealistic to develop an empirical model that is universally applicable to the eutrophication management of a broad spectrum of lake ecosystems worldwide, our objective was to develop a robust statistical approach to dealing with ecosystem-specific data that could be utilised for management entities with access to similar monitoring data sets. Indeed, it was difficult to develop ecosystem-specific eutrophication models for individual lakes in the 1960s when a multiple-lake modelling approach was developed given the scarcity of water quality monitoring data. The availability of water quality observations,

however, has since increased for many large lakes (e.g. Great lakes, Lake Champlain, Lake Okeechobee, U.S.A.; Three Gorges Reservoir, Lake Taihu, Lake Caohu, China). Thus, it is time for lake researchers and managers to leverage these large data sets to develop and apply statistical models tailored to individual lakes when possible. In addition, advances in modern statistical methods (e.g. quantile regression) and corresponding software (e.g. R quantreg package) over the past 50 years make it easier to analyse these growing water quality data sets to construct more sophisticated and ecologically meaningful eutrophication models.

Additionally, while in Lake Champlain, the PS-QR-coupled approach suggests that phosphorus reduction is critical for eutrophication control, some of the novelty of this approach from a theoretical standpoint improves the understanding of the limiting factors, or their co-limitations, of such ecological processes. Indeed, co-limitation is becoming a more prevalent paradigm in lake ecology, particularly in highly eutrophic systems (Elser *et al.*, 2007), and statistical approaches that more accurately quantify the relationship between both nutrients and chlorophyll are critical to understanding the drivers of ecological processes in such systems. As we reveal the important role of reduced phosphorus on nitrogen limitation using the PS-QR-coupled approach, this approach may also be useful for understanding the interacting roles of multiple limiting nutrients and predicting reduction targets that consider changes in other controlling factors (e.g. flushing, grazing, rainfall, temperature) when the corresponding parameters are available for researchers. Expanding the PS-QR-coupled approach to include other parameters could actually prove a powerful way forward for more complete theoretical knowledge of lake-specific resource limitation needed by the management community. Furthermore, this approach allows the research community to investigate these important ecological processes using existing long-term data sets, which could then be used for stand-alone research projects or to help guide additional ecological experiments or sampling strategies. As such, the approach developed in this study is broadly applicable to research and management communities interested in developing ecologically meaningful and empirically robust models for the nutrient reduction targets of individual lake-specific ecosystems contaminated or threatened by nutrient loading.

Acknowledgments

Support was provided by Vermont EPSCoR with funds from the National Science Foundation Grant EPS-1101317.

The authors are grateful to the staff of the Vermont Department of Environmental Conservation and New York State Department of Environmental Conservation, who collected and processed data for a long-term monitoring programme (LTMP, 1992–2012) on Lake Champlain. We appreciate J. Jones, J. Stockwell, C. Giles and T. Gearhart for their constructive comments in improving the early version of this manuscript and R. Koenker for assistance with quantile regression as well as three anonymous reviewers and associate editor (Prof. Roger Jones) for their productive comments in revising this manuscript. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. [Correction added on 5 February 2016, after first online publication on 9 June 2015: In the acknowledgement section the Grant number has been amended to 1101317].

References

- Abell J., Özkundakci D., Hamilton D. & Jones J. (2012) Latitudinal variation in nutrient stoichiometry and chlorophyll-nutrient relationships in lakes: a global study. *Fundamental and Applied Limnology*, **181**, 1–14.
- Bryhn A.C. & Håkanson L. (2009) Eutrophication: model before acting. *Science*, **324**, 723.
- Cade B.S. & Guo Q. (2000) Estimating effects of constraints on plant performance with regression quantiles. *Oikos*, **91**, 245–254.
- Cade B.S. & Noon B.R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, **1**, 412–420.
- Cade B.S., Noon B.R. & Flather C.H. (2005) Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology*, **86**, 786–800.
- Cade B.S., Terrell J.W. & Schroeder R.L. (1999) Estimating effects of limiting factors with regression quantiles. *Ecology*, **80**, 311–323.
- Cardoso A.C., Solimini A., Premazzi G., Carvalho L., Lyche A. & Rekolainen S. (2007) Phosphorus reference concentrations in European lakes. *Hydrobiologia*, **584**, 3–12.
- Carvalho L., McDonald C., de Hoyos C., Mischke U., Phillips G., Borics G. *et al.* (2013) Sustaining recreational quality of European lakes: minimizing the health risks from algal blooms through phosphorus control. *Journal of Applied Ecology*, **50**, 315–323.
- Cha Y., Stow C.A. & Bernhardt E.S. (2013) Impacts of dreissenid mussel invasions on chlorophyll and total phosphorus in 25 lakes in the USA. *Freshwater Biology*, **58**, 192–206.
- Chen C., Kamman N., Williams J., Bugge D., Taylor V., Jackson B. *et al.* (2012) Spatial and temporal variation in mercury bioaccumulation by zooplankton in Lake Champlain (North America). *Environmental Pollution*, **161**, 343–349.
- Conley D.J., Paerl H.W., Howarth R.W., Boesch D.F., Seitzinger S.P., Havens K.E. *et al.* (2009a) Controlling eutrophication: nitrogen and phosphorus. *Science*, **323**, 1014–1015.
- Conley D.J., Paerl H.W., Howarth R.W., Boesch D.F., Seitzinger S.P., Havens K.E. *et al.* (2009b) Eutrophication: time to adjust expectations–response. *Science*, **324**, 724–725.
- Dillon P.J. & Rigler F.H. (1974) The phosphorus-chlorophyll relationship in lakes. *Limnology and Oceanography*, **19**, 767–773.
- Dodds W.K., Bouska W.W., Eitzmann J.L., Pilger T.J., Pitts K.L., Riley A.J. *et al.* (2009) Eutrophication of U.S. freshwaters: analysis of potential economic damages. *Environmental Science & Technology*, **43**, 12–19.
- Elser J.J., Bracken M.E.S., Cleland E.E., Gruner D.S., Harpole W.S., Hillebrand H. *et al.* (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology letters*, **10**, 1135–1142.
- Guilbert J., Beckage B., Winter J.M., Horton R.M., Perkins T. & Bombles A. (2014) Impacts of projected climate change over the Lake Champlain Basin in Vermont. *Journal of Applied Meteorology and Climatology*, **53**, 1861–1875.
- Håkanson L., Blenckner T., Bryhn A.C. & Hellström S.S. (2005) The influence of calcium on the chlorophyll–phosphorus relationship and lake Secchi depths. *Hydrobiologia*, **537**, 111–123.
- Higgins S.N., Vander Zanden M.J., Joppa L.N. & Vadeboncoeur Y. (2011) The effect of dreissenid invasions on chlorophyll and the chlorophyll: total phosphorus ratio in north-temperate lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **68**, 319–329.
- Huo S., Xi B., Ma C. & Liu H. (2013) Stressor-response models: a practical application for the development of lake nutrient criteria in China. *Environmental Science & Technology*, **47**, 11922–11923.
- Huszar V.L.M., Caraco N.F., Roland F. & Cole J. (2006) Nutrient-chlorophyll relationships in tropical subtropical lakes: do temperate models fit? *Biogeochemistry*, **79**, 239–250.
- Jacoby C.A. & Frazer T.K. (2009) Eutrophication: time to adjust expectations. *Science*, **324**, 723.
- Jones J.R., Knowlton M.F. & Kaiser M.S. (1998) Effects of aggregation on chlorophyll-phosphorus relations in Missouri reservoirs. *Lake and Reservoir Management*, **14**, 1–9.
- Jones J.R., Obrecht D.V. & Thorpe A.P. (2011) Chlorophyll maxima and chlorophyll:total phosphorus ratios in Missouri reservoirs. *Lake and Reservoir Management*, **27**, 321–328.
- Kaiser M.S., Speckman P.L. & Jones J.R. (1994) Statistical models for limiting nutrient relations in inland waters. *Journal of the American Statistical Association*, **89**, 410–423.
- Koenker R. & Bassett G. Jr (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker R. & Machado J.A. (1999) Goodness of fit and related inference processes for quantile regression.

- Journal of the American Statistical Association*, **94**, 1296–1310.
- Koenker R., Portnoy S., Tian Ng T., Zeileis A., Grosjean P. & Ripley B.D. (2013) quantreg: Quantile Regression and related models. R package version 5.05. <http://CRAN.R-project.org/package=quantreg>
- Levine S.N., Lini A., Ostrofsky M.L., Bunting L., Burgess H., Leavitt P.R. *et al.* (2012) The eutrophication of Lake Champlain's northeastern arm: insights from paleolimnological analyses. *Journal of Great Lakes Research*, **38**, 35–48.
- Malve O. & Qian S.S. (2006) Estimating nutrients and chlorophyll a relationships in Finnish Lakes. *Environmental Science & Technology*, **40**, 7848–7853.
- Marsden J.E. & Langdon R.W. (2012) The history and future of Lake Champlain's fishes and fisheries. *Journal of Great Lakes Research*, **38**, 19–34.
- McKean J.W. & Sievers G.L. (1987) Coefficients of determination for least absolute deviation analysis. *Statistics & Probability Letters*, **5**, 49–54.
- Mooij W.M., Trolle D., Jeppesen E., Arhonditsis G., Belolipetsky P.V., Chitamwebwa D.B.R. *et al.* (2010) Challenges and opportunities for integrating lake ecosystem modeling approaches. *Aquatic Ecology*, **44**, 633–667.
- O'Donnell D.M., Effler S.W., Perkins M. & Strait C. (2013) Optical characterization of Lake Champlain: spatial heterogeneity and closure. *Journal of Great Lakes Research*, **39**, 247–258.
- Paerl H.W. & Scott J.T. (2010) Throwing fuel on the fire: synergistic effects of excessive nitrogen inputs and global warming on harmful algal blooms. *Environmental Science & Technology*, **44**, 7756–7758.
- Phillips G., Pietiläinen O.P., Carvalho L., Solimini A., Solheim L.A. & Cardoso A. (2008) Chlorophyll–nutrient relationships of different lake types using a large European dataset. *Aquatic Ecology*, **42**, 213–226.
- Prairie Y.T., Duarte C.M. & Kalff J. (1989) Unifying nutrient–chlorophyll relationships in lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 1176–1182.
- R Development Core Team. (2014) *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Sakamoto M. (1966) Primary production by phytoplankton community in some Japanese lakes and its dependence on lake depth. *Archiv Fur Hydrobiologie*, **62**, 1–28.
- Schelske C.L. (2009) Eutrophication: focus on phosphorus. *Science*, **324**, 722.
- Schindler D.W. (2006) Recent advances in the understanding and management of eutrophication. *Limnology and Oceanography*, **51**, 356–363.
- Schindler D.W. & Hecky R.E. (2009) Eutrophication: more nitrogen data needed. *Science*, **324**, 721–722.
- Schooley R.L. & Wiens J.A. (2005) Spatial ecology of cactus bugs: area constraints and patch connectivity. *Ecology*, **86**, 1627–1639.
- Smeltzer E., Shambaugh A.D. & Stangel P. (2012) Environmental change in Lake Champlain revealed by long-term monitoring. *Journal of Great Lakes Research*, **38**, 6–18.
- Smith V.H., Joye S.B. & Howarth R.W. (2006) Eutrophication of freshwater and marine ecosystems. *Limnology and Oceanography*, **51**, 351–355.
- Smith V.H. & Schindler D.W. (2009) Eutrophication science: where do we go from here. *Trends in Ecology and Evolution*, **24**, 201–207.
- Smith V.H. & Shapiro J. (1981) Chlorophyll–phosphorus relations in individual lakes - Their importance to lake restoration strategies. *Environmental Science & Technology*, **15**, 444–451.
- Stauffer R.E. (1991) Environmental factors influencing chlorophyll nutrient relationships in Lakes. *Freshwater Biology*, **25**, 279–295.
- Sterck F., Markesteijn L., Toledo M., Schieving F. & Poorter L. (2014) Sapling performance along resource gradients drives tree species distributions within and across tropical forests. *Ecology*, **95**, 2514–2525.
- Stow C.A. & Cha Y. (2013) Are chlorophyll *a*–total phosphorus correlations useful for inference and prediction? *Environmental Science & Technology*, **47**, 3768–3773.
- Stow C.A., Reckhow K.H. & Qian S.S. (2006) A Bayesian approach to retransformation bias in transformed regression. *Ecology*, **87**, 1472–1477.
- U. S. EPA (United States Environmental Protection Agency). (2000) Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs. http://www2.epa.gov/sites/production/files/documents/guidance_lakes.pdf
- VTDEC (Vermont Department of Environmental Conservation) & NYDEC (New York State Department of Environmental Conservation). (2013) Long-term water quality and biological monitoring project for Lake Champlain–Quality Assurance Project Plan. Grand Isle, Vermont. http://www.watershedmanagement.vt.gov/lakes/htm/lp_longterm.htm
- Wagner C.E., Harmon L.J. & Seehausen O. (2014) Cichlid species–area relationships are shaped by adaptive radiations that scale with area. *Ecology Letters*, **17**, 583–592.
- Wagner T., Soranno P.A., Webster K.E. & Cheruvilil K.S. (2011) Landscape drivers of regional variation in the relationship between total phosphorus and chlorophyll in lakes. *Freshwater Biology*, **56**, 1811–1824.
- Wang H., Liang X., Jiang P., Wang J., Wu S. & Wang H. (2008) TN: TP ratio and planktivorous fish do not affect nutrient–chlorophyll relationships in shallow lakes. *Freshwater Biology*, **53**, 935–944.
- Wang H., Wang H., Liang X. & Wu S. (2014) Total phosphorus thresholds for regime shifts are nearly equal in subtropical and temperate shallow lakes with moderate depths and areas. *Freshwater Biology*, **59**, 1659–1671.

(Manuscript accepted 14 May 2015)