

# Reconstructing phosphorus levels using models based on the modern diatom assemblages of 55 lakes in southern Quebec

Roxane Tremblay, Reinhard Pienitz, and Pierre Legendre

**Abstract:** This study shows how modern diatom assemblages, obtained from the surface sediments of 55 deep and shallow lakes in southern Quebec and representing a gradient in trophic state and specific conductivity, can be used to reconstruct the historical water column concentrations of total phosphorus (TP) over time. Inference models were developed with the aid of regression analyses based on weighted means. The TP models demonstrate statistical performances equal to or superior to TP models developed in the past for other regions in Canada and elsewhere in the world. A reconstruction of TP values, based on fossil diatom assemblages of Lac Nairne (Charlevoix, QC), was conducted. Comparing the TP inferred values with the measured values and those estimated using export coefficients, the reliability of the model was proven. In addition, the comparison of the results obtained from reconstructions performed with two inference models (Quebec and Ontario lakes) also revealed that models predicted TP and not a collinear variable.

**Résumé :** Cet article vise à montrer que les assemblages de diatomées modernes répertoriées dans les sédiments de surface de 55 lacs du Québec méridional, représentatifs d'un gradient d'états trophiques, permettent de reconstituer les teneurs historiques en phosphore total (PT) qui prévalaient dans la colonne d'eau des lacs du Québec. Pour ce faire, un modèle d'inférence a été développé à l'aide d'analyses de régressions basées sur les moyennes pondérées. Les performances statistiques du modèle de PT sont équivalentes ou supérieures aux modèles développés pour des régions différentes au Canada ou ailleurs dans le monde. Une reconstitution des valeurs de PT, basée sur les assemblages de diatomées fossiles du lac Nairne (Charlevoix, QC) a été menée. La comparaison des valeurs de PT inférées avec les valeurs mesurées ainsi que celles estimées à l'aide des coefficients d'exportation a montré la justesse du modèle. De plus, la comparaison des résultats provenant de reconstitutions menées avec deux modèles d'inférence (lacs du Québec et de l'Ontario) a également montré que les modèles prédisent le PT et non une variable colinéaire.

## Introduction

In southern Quebec, numerous watersheds are under increasing pressure from urban, agricultural, and resort activities. These activities accelerate natural eutrophication and the degradation of water quality. The severity of the impacts caused by chemicals on water bodies depends primarily on their role in disturbing or maintaining water quality and aquatic ecosystems, as well as their influence on primary production or biomass. The element that typically limits the primary productivity of freshwater environments is the availability of phosphorus (Schindler 1974).

The objective of this study was to develop inference models to reconstruct total phosphorus (TP) values in lake waters, based on ecological preferences of diatoms (class Bacillariophyceae) obtained from the surface sediments of 56 lakes located in southern Quebec. Diatoms are microscopic algae that live suspended in the water column or attached to living or inert substrates. They are particularly sensitive to water physics and chemistry, including specific conductivity, pH, nutrient concentrations (phosphorus and nitrogen), and organic matter contents. Research conducted by van Dam et al. (1994) and Smol and Stoermer (2010) reviewed the impact of these variables on the structure and composition of diatom assemblages.

The preservation of diatom frustules in lake sediments allows the development of inference models for the reconstruction of environmental variables based on species–environment relation-

ships and the determination of ecological preferences of various diatom taxa. These models are mostly developed using canonical ordination techniques (ter Braak 1985, 1986) and regression analysis based on weighted means (e.g., ter Braak and van Dam 1989). The methodology for developing these types of inference models is well documented (e.g., Birks 1995, 1998; Birks et al. 2012). To date, many inference models have been developed throughout the world for different countries and variables (see Smol and Stoermer 2010). However, the usefulness of any given model is generally restricted to its region of origin and not all models are without problems (reviewed in Telford and Birks 2009; Juggins 2013a).

Briefly, increasing concentrations of phosphorus in lakes generally lead to increased primary productivity and biomass of plants and algae that result in adverse effects, such as loss of diversity, increase of tolerant species (e.g., cyanobacteria), and decrease in water transparency. These conditions are caused by increases in biomass and organic materials that cause sedimentation rates to increase and, finally, result in oxygen deficiency in the hypolimnion (Mason 1991; Horne and Goldman 1994). Consequently, the study of a set of lakes with a large trophic state gradient cannot be properly done without violating one of the assumptions usually associated with the development of inference models, namely that “environmental variables other than the one of interest have negligible influence, or their joint distribution with the environmental variable does not change with time” (Juggins 2013a). Furthermore, we believe that the “rule of

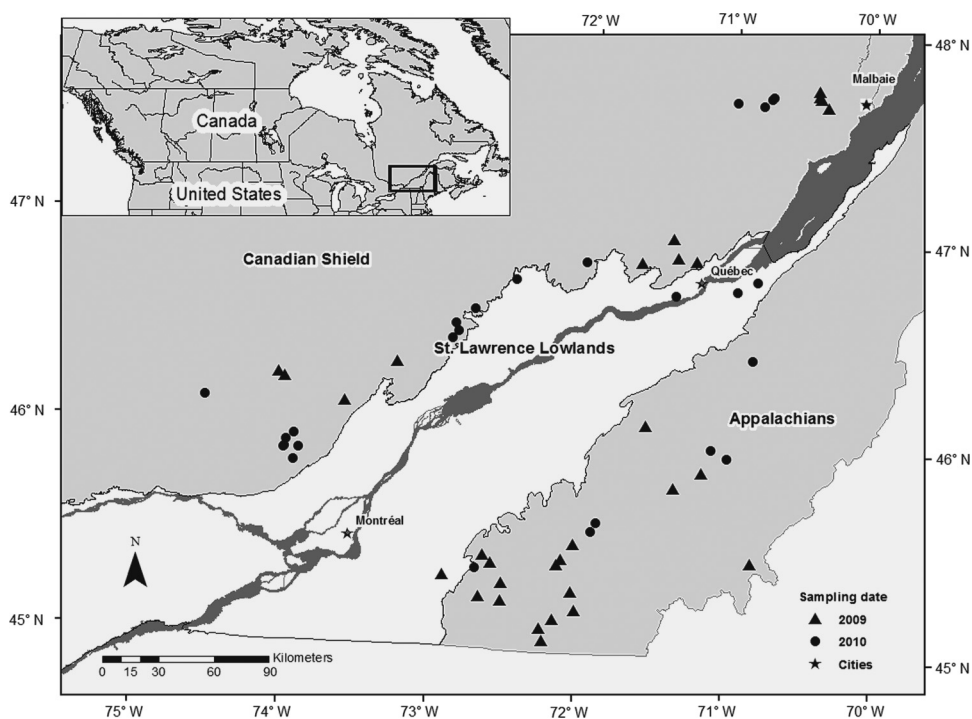
Received 29 August 2013. Accepted 5 February 2014.

R. Tremblay\* and R. Pienitz. Centre d'études nordiques, Pavillon Abitibi-Price, Université Laval, 2405, rue de la Terrasse, Québec, QC G1V 0A6, Canada. P. Legendre. Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succursale, Centre-ville, Montréal, QC H3C 3J7, Canada.

**Corresponding authors:** Roxane Tremblay (e-mail: roxane.tremblay@cima.ca), Reinhard Pienitz (e-mail: reinhard.Pienitz@cen.ulaval.ca).

\*Present address: Environmental Project Leader, CIMA\*, 2030 de la Rive-Sud Blvd., Suite 201, Lévis, QC G6W 2S6, Canada.

**Fig. 1.** Map of the study area showing the locations of the 56 lakes and the geological regions of the southern Laurentian Mountains, the Saint Lawrence Lowlands, and the Appalachians.



thumb” proposed by [ter Braak \(1988\)](#), which states that reconstructing a variable is possible if the variance of the diatom distribution explained by this variable in a canonical ordination is greater than the proportion of variance of the first axis of an unconstrained ordination, is not binding to obtain a robust inference model. Of course, this rule makes ecological sense. It ensures that the diatom gradient is best explained linearly when forced to be organized according to the variable of interest. However, inference models can be efficient without necessarily following this rule (e.g. [Hall and Smol 1992](#); [Werner and Smol 2005](#); [Yang et al. 2008](#)).

The inference model developed here for TP concentrations will help determine (i) the natural background found in southern Quebec lakes, (ii) the magnitude of human-induced changes in the watershed following the arrival of Europeans in the early 17th century, and (iii) the ecosystem thresholds that exist to maintain or re-establish equilibrium states in the impacted lakes. Finally, this paper shows that diatom-based models obtained from the surface sediment with modern assemblages of 56 lakes located in southern Quebec, and representing a large trophic gradient, allow the reconstruction of historical levels of TP in the water column of southern Quebec lakes.

## Materials and methods

### Study area and available data

In total, 56 lakes were selected across different geological regions of southern Quebec: the southern Laurentian Mountains (A), and Saint Lawrence Lowlands (SL) ([Li and Ducruc 1999](#)) (Fig. 1; Table A1). The southern Laurentian Mountains cover part of the Southwest Canadian Shield in Quebec. The bedrock consists primarily of Precambrian granite and gneisses covered by thin glacial deposits of the last Wisconsinan glaciation. In the Appalachian region, sedimentary and sometimes volcanic bedrock is covered by glacial deposits that range from thin to thick. The Saint Lawrence Lowlands consist of a platform of marine deposits originating from the postglacial

Champlain Sea, as well as glacial and alluvial deposits and peatlands.

In most cases, the catchments surrounding the lakes are urbanized, yet to different degrees. The [Strahler \(1957\)](#) hydrological classification system was used to classify lakes from 0 to 6 for their positions in basins; the value zero was assigned to the headwater lakes, 1 to lakes receiving waters from a first order tributary, and so on.

The lakes selected from the three regions are distributed along a trophic gradient that ranges from ultra-oligotrophic to hyper-eutrophic conditions. The mean of concentrations ( $n = 3$ ) of TP varied between 0.9 and 173.3  $\mu\text{g}\cdot\text{L}^{-1}$ , the concentrations of chlorophyll *a* (Chl<sub>a</sub>) varied between 0.4 and 95.6  $\mu\text{g}\cdot\text{L}^{-1}$ , and water transparency between 0.4 and 25.0 m Secchi depth. Specific conductivity varied between 11 and 712  $\mu\text{S}\cdot\text{cm}^{-1}$ . The alkalinity and epilimnetic pH varied between 1.8 and 114.7  $\text{mg}\cdot\text{L}^{-1}$   $\text{CaCO}_3$  and 6.4 and 9.5, respectively. In addition, the lakes themselves were highly variable both limnologically and morphometrically (Table A2). The lakes were either polymictic (21 lakes) or dimictic (35 lakes).

### Lake sampling and sample preparation

#### Diatoms

In total, 56 lakes were sampled between early June and the end of September, which was the stratification period in the dimictic lakes. Each lake was visited three times during the summers of either 2009 or 2010, with sediments obtained from the deepest part of the lake or from one of its secondary basins using an Aquatic Research Instruments percussion corer. Modern diatom assemblages were extracted from the sediments in the Aquatic Paleoecology Laboratory at Université Laval following the method of [Pienitz et al. \(1995\)](#). In total, 400 diatom valves were identified for each level core. The main taxonomic keys used for their identification were [Krammer and Lange-Bertalot \(1986, 1988, 1991a, 1991b\)](#), [Fallu et al. \(2000\)](#), [Camburn and Charles \(2000\)](#), [Lavoie et al. \(2008\)](#), and [Siver and Kling \(1997\)](#).

**Table 1.** Environmental variables retained after applying the step-by-step selection criteria on the four types of variables.

Matrix	No.	Variable	R <sup>2</sup> <sub>adjusted</sub> cumulative	AIC*	F	p value of partial tests (199 perm.)
Causative variables (chemical or physical)	1	Total phosphorus (TP)	0.0708	-33.77	5.19	0.005
	2	Aluminium (Al)	0.1088	-35.15	3.30	0.005
	3	Iron (Fe)	0.1241	-35.18	1.92	0.010
	4	Temperature (Temp.)	0.1377	-35.15	1.82	0.005
	5	Sodium (Na)	0.1522	-35.21	1.87	0.005
	6	Alkalinity (Alk.)	0.1685	-35.43	1.98	0.005
	7	NiT:TP ratio	0.1824	-35.53	1.83	0.005
	8	Silica (Si)	0.1942	-35.52	1.70	0.015
	9	Conductivity (Cond.) R <sup>2</sup> for all 20 variables	0.2042 0.2181	-35.42	1.58	0.015
Response variables (biological, chemical, or physical)	10	Epilimnic pH (pH)	0.0692	-33.67	5.09	0.005
	11	Depth of photic zone (Photic)	0.1092	-35.18	3.42	0.005
	12	Transparency (Secchi)	0.1335	-35.79	2.48	0.005
		R <sup>2</sup> for all 7 variables	0.1453			
Morphometric and geographical variables	13	Strahler class (Strahler)	0.0794	-29.73	1.79	0.005
	14	Maximal depth (ZStation)	0.1364	-32.46	4.23	0.005
	15	Watershed on lake area (W/L_Area)	0.1474	-32.36	1.62	0.025
		R <sup>2</sup> for all 15 variables	0.1638			
Geological variables	16	Ecoregion (Ecoreg.)	0.0457	-31.32	2.32	0.005
	17	Carbonate rocks (CARBO)	0.0591	-31.18	1.75	0.020
	18	Gneiss rocks (GNEISS)	0.0724	-31.06	1.75	0.020
	19	Lacustrine deposits (LACU)	0.0876	-31.10	1.85	0.005
		R <sup>2</sup> for all 15 variables	0.0934			

\*Akaike information criterion.

### Environmental variables

In total, 57 variables were studied to determine their influence on the composition of the diatom assemblages. These variables were divided into four groups numbered 1 to 4: 1, causative variables (chemical or physical) (20); 2, response variables (biological, chemical, or physical) (7); 3, morphometric and geographical variables (15); and 4, geological variables (15). Groups are shown in the Pearson correlation matrix (Table A4).

Sampling of water and physical and chemical profiling were conducted at each visit. A Quanta sensor was used to measure pH, specific conductivity, temperature, and dissolved oxygen at 1 m intervals. Values represent the average of three results ( $n = 3$ ), with measurements completed during the summer. Water transparency was evaluated using a 20 cm diameter Secchi disc. Each of the values for temperature, pH, specific conductivity, and dissolved oxygen in the photic zone were compiled to establish the average summertime depth of the photic zone. Photic depth was determined by multiplying the Secchi depth by a factor of 1.7 for tea-colored lakes, or 2.4 for clear or blue-colored lakes, the latter being located on granitic and gneissic bedrock and possessing lower levels of dissolved organic carbon. These factors are based on studies by Idso and Gilbert (1974), Koenings and Edmundson (1991), and Kalff (2002).

### Laboratory water chemistry

Some of the analyses were conducted at INRS-ETE laboratories (Québec, Que.). Inductively coupled plasma atomic emission spectroscopy (ICP-AES Varian Vista AX) determined metal levels, ion chromatography (Dionex) measured anion concentrations, whereas a total organic carbon (TOC) analyzer (Shimadzu TOC) measured dissolved organic carbon (DOC) and total inorganic nitrogen (TiN) concentrations. The analysis of TP (dissolved and particulate) was conducted by Centre d'expertise en analyse environnementale du Québec (CEAEQ) using the MA. 303 - P 5 method (CEAEQ 2008). Chla was quantified by fluorometry at INRS-ETE, while the total suspended solids (TSS) was determined in the Aquatic Paleoecology Laboratory at Université Laval following filtration and drying at 105 °C.

Values for metals, anions, TP, TiN, and DOC were obtained from an integrated sample from the surface to 1 m depth. Finally, the TSS value is the result of an integrated sample obtained from the surface to 4 m depth, or to 0.5 m from the sediment surface for lakes that were less than 4 m deep.

### Morphometric and geological variables

Lake volume was determined using bathymetric curves or extrapolated using the method of Hollister and Milstead (2010), with the aid of the computer program ArcGIS from ESRI. This program was also used to determine the geological and surficial deposit characteristics for each watershed.

### Statistical analysis

All statistical analyses were conducted using R software (R Development Core Team 2013). The selection of explanatory environmental variables was based on canonical redundancy analysis (RDA). Because of the long ecological gradient causing the appearance of many zeros in the data, transformation-based RDA (tb-RDA, using the Hellinger transformation) was used for modelling the species–environment relationships (Legendre and Birks 2012, Legendre and Legendre 2012). The species–environment analysis using tb-RDA produced a higher R<sup>2</sup> than CCA (R<sup>2</sup> tb-RDA = 0.53 and R<sup>2</sup> CCA = 0.47). Note that the diatom matrix was not transformed before CCA. The two canonical ordinations were performed with all taxa and the 19 environmental variables listed in Table 1. The distribution of sites and taxa in the RDA and the CCA was comparable, except for very different outlier assemblages. These samples had more of an “eccentric” position (far from the origin of the graph) in the CCA than in the RDA. This situation is explained by the Hellinger transformation applied to the data before computing RDA. The development of inference models was based on weighted average regression analysis (WA and WA-PLS).

### Transformation of environmental variables

The environmental variables studied are for the most part quantitative, except for those related to different types of mixing (Type) and geological regions (Ecoreg.), which are categorical. To reduce

the skewness of the frequency distributions and increase the linearity of the relationships, some variables were  $\log_{10}$  or square-root transformed (Table A3). Pearson correlation tests determined the degree of correlation among the environmental variables. Because of the high number of simultaneous comparisons, a Bonferroni correction was applied to the  $p$  values to obtain adjusted probabilities.

#### Transformation of the diatom taxa matrix

The Hellinger transformation was applied to the assemblage data (Legendre and Gallagher 2001). This transformation forces the RDA to preserve the Hellinger distance, which is a distance appropriate for the study of community composition data (Legendre and Legendre 2012) and beta diversity (Legendre and De Cáceres 2013). That distance is insensitive to double zeros and reduces the importance of extremely abundant taxa, reducing the effect of exponential growth where favorable environmental conditions encourage it. For example, it reduced the extreme dominance of taxa produced during blooms, as in the case of *Asterionella formosa* and *Fragilaria crotonensis*. Unlike canonical correspondence analysis (CCA), RDA does not require elimination of rare taxa before analysis.

#### Selection of environmental variables

Selection of explanatory variables by RDA was conducted to determine which environmental variables significantly explained the distribution of the diatom assemblages throughout the lakes. Step-by-step selection, following the method described by Blanchet et al. (2008), was conducted separately for the four types of variables: (i) causative variables (chemical or physical), (ii) response variables (biological, physical, or chemical; e.g., Chl $a$ ), (iii) morphometric and geographic variables, and (iv) geological variables. The goal was to retain only the variables that explained a complementary and non-redundant part of the diatom among-lake variation for each variable type. This analysis was conducted using the ordiR2step() function of the VEGAN software package (Oksanen et al. 2013). After selection of the variables in each variable type class, variance partitioning (Borcard et al. 1992) was conducted by RDA between the response matrix (diatom assemblages) and the four environmental (explanatory) matrices, to determine their unique and joint influences on the variability of the assemblages. This analysis was conducted using the varpart() function of the VEGAN package.

A final tb-RDA was performed using the retained environmental variables to illustrate the relationship between the significant explanatory variables and their influence on the diatom assemblages. The significance of all terms together has been assessed using permutation tests with the ANOVA function of the VEGAN package.

#### Inference models

The development of the inference models initially required the exclusion of lakes that were outliers with respect to the taxonomic and water physical and chemical data. The justifications are explained in the Results and discussion.

Inference models were developed to predict the TP concentrations using weighted averaging regressions (WA) and weighted averaging partial least squares regressions (WA-PLS). The two methods are based on the assumption that a species is most abundant when conditions are optimal for its growth (ter Braak and van Dam 1989). The WA function supports both classic and inverse deshrinking regression. Only the classic WA method for unimodal ecological patterns was used, as the inverse WA gives the same result as a one-component WA-PLS. WA-PLS uses the residuals of the model by integrating them as components in a successive manner (ter Braak and Juggins 1993). This method is thus considered more powerful than WA, but WA-PLS reveals also two disadvantages. First, it can easily over-fit the data, and second,

it needs careful model diagnosis to understand why higher components improve the fit to the training data set (Juggins and Birks 2012).

The weighted averaging regressions and the development of the inference models were conducted using the wa() and waps() functions of the PALTRAN software package in R (Adler 2010), as well as the WA() and WAPLS() functions of the rioja software package (Juggins 2013b).

#### Evaluation of model performance

In a first step, two performance criteria were used to evaluate the inference model: the coefficient of determination ( $R^2$ ) and the value of the square root of the sum of the squared errors (RMSE). In a second step, a model validation was conducted to determine if the model was too specific to the data for which it was developed (over calibrated). Indeed, a model is considered too specific when it has problems in adequately predicting samples that do not belong to the data set used for its calibration. A generalist model will usually provide better results. Two nonparametric validation tests were used to judge the performance of the models: bootstrap and cross-validation (leave-one-out or jackknife). The performance of the validation tests is given by the validated coefficient of determination ( $R^2_{\text{boot or jack}}$ ) and the value of the square root of the sum of the predicted squared errors (RMSEP). The retained model displayed the smallest RMSEP and the smallest maximum error. In a third step, the potential influence of spatial autocorrelation was verified. Telford and Birks (2005) argue that spatial structure in environmental variables causes cross-validated estimates of the predictive power of transfer functions to be over-optimistic, and can result in inappropriate model selection. Thus, we assessed whether the effect of spatial autocorrelation could have an undesired influence in our model. To this effect, we followed the proposed Telford and Birks (2009) procedure. This implies the calculation of the effect of site elimination for different geographical distances in a validation test by permutation. Then, the performance ( $R^2$ ) of these models is compared to that of the models that randomly eliminate a portion of the sites. This analysis was performed with the rn function in the palaeoSig package (Telford 2013). In a fourth step, taxa occurrences were also used to judge the performance of the inference models, with this criterion also being evaluated in terms of the value of the maximum error (or bias), which should be minimized. In a fifth step, the residuals were examined as a function of the model predictions, as recommended by Racca and Prairie (2004), to document the bias that can be attributed to each model.

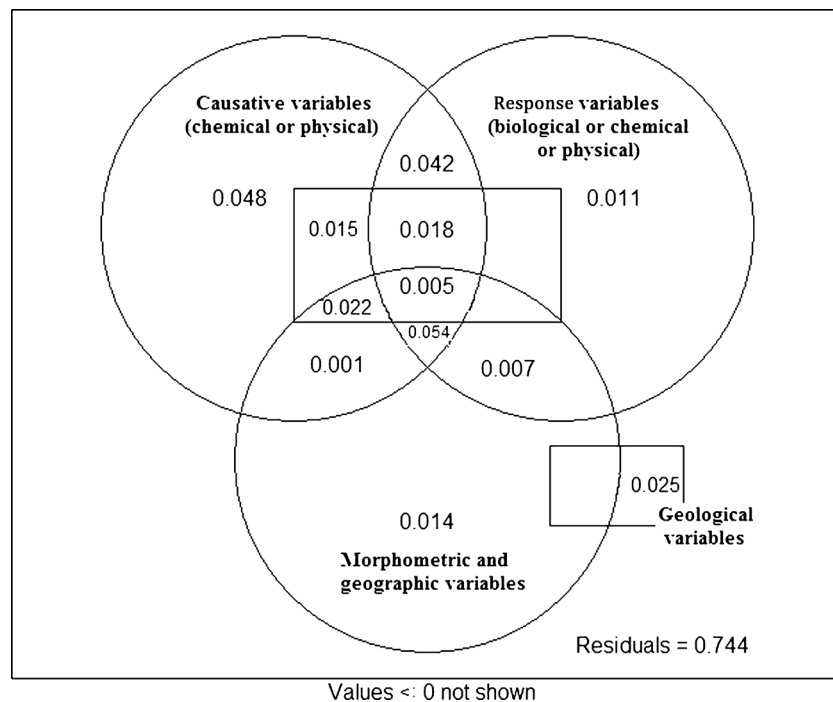
#### Analogy test

An analogy test was conducted between the diatom assemblages predicted by the inference model and those found in the reconstructed sedimentary sequence to document the similarities between the assemblages of the modern data set and those of the sedimentary sequence. The method followed Overpeck et al. (1985). The dissimilarity between the fossil assemblages and those of the inference model was compared using Orłóci's chord distance. If the dissimilarity of the fossil assemblage exceeds the 95th percentile of the dissimilarity of the model's assemblages, it is not appropriate to use the inference model on the fossil assemblages. The representation of the dominant species within the data set was also an important criterion to determine the appropriateness of the model used.

#### Data used for testing the inference models

The TP inference models were tested using data that were acquired in two previous studies conducted on Lac Nairne and its watershed. Lac Nairne (47°41'N, 70°21'W) is located within the municipality of Saint-Aimé-des-Lacs in Charlevoix, on the northern shore of the Saint Lawrence River, approximately 125 km east of the city of Québec. The lake has an area of 2.43 km<sup>2</sup> with a

**Fig. 2.** Venn diagram of the partitioning of the response data (diatoms) explained by four data sets: 1, causative variables (chemical or physical) (20); 2, response variables (biological, chemical, or physical) (7); 3, morphometric and geographical variables (15); and 4, geological variables (15).



watershed of 28.95 km<sup>2</sup>. Land use and the geology of the watershed, including the lake limnological characteristics, are described in Labrecque et al. (2012).

The study of Roy (2012) included two sediment core sequences obtained from Lac Nairne (N-01), obtained during the summer of 2009. The N-01 sequence is 50 cm long and was used to identify and count the fossil diatom assemblages. Using the fossil diatom assemblages of Lac Nairne, a semiquantitative index of trophic state was calculated with the aid of the main diatom taxa identified and counted. The index was constructed by documenting the trophic preferences of the main taxa (arbitrary values forming a linear scale) and multiplying this trophic "rating" by the taxon abundances and then summing the values for all taxa in the assemblage. Thus, to develop the trophic state index, a species with oligotrophic preference was assigned a lower rating than a species with eutrophic preference. The final index contained 27 taxa comprised of the most abundant taxa with known autecologies (Roy 2012). This rating system was developed to allow the comparison of the performance of an interpreted change profile sequence with a semiquantitative index, as well as against an inference model. Considering that the inference models developed in this study were based on the ecological principle that a taxon shows maximum growth in an environment with optimal conditions, it is reasonable to expect similar trends between the profile of the semiquantitative trophic index and the profile of inferred TP values, if phosphorus is the limiting nutrient. A Pearson correlation test was conducted between the inferred TP concentration values and those obtained by the semiquantitative index to compare them. Data relating to the phosphorus budget of the Lac Nairne watershed (Labrecque 2013), according to current and past land uses, were used to compare the values inferred by our TP concentration models and those estimated by Labrecque (2013). The latter used phosphorus export coefficients, based on the various land uses (forestry, agriculture, grassland, urbanization, etc.) and on point sources (e.g., septic tanks) to calculate TP inputs into the lake over several years. Thus, the results correspond to TP concentration in micrograms per litre ( $\mu\text{g}\cdot\text{L}^{-1}$ ) transported to the lake.

The historical land use of the region was interpreted through analysis of aerial photographs (Labrecque 2013).

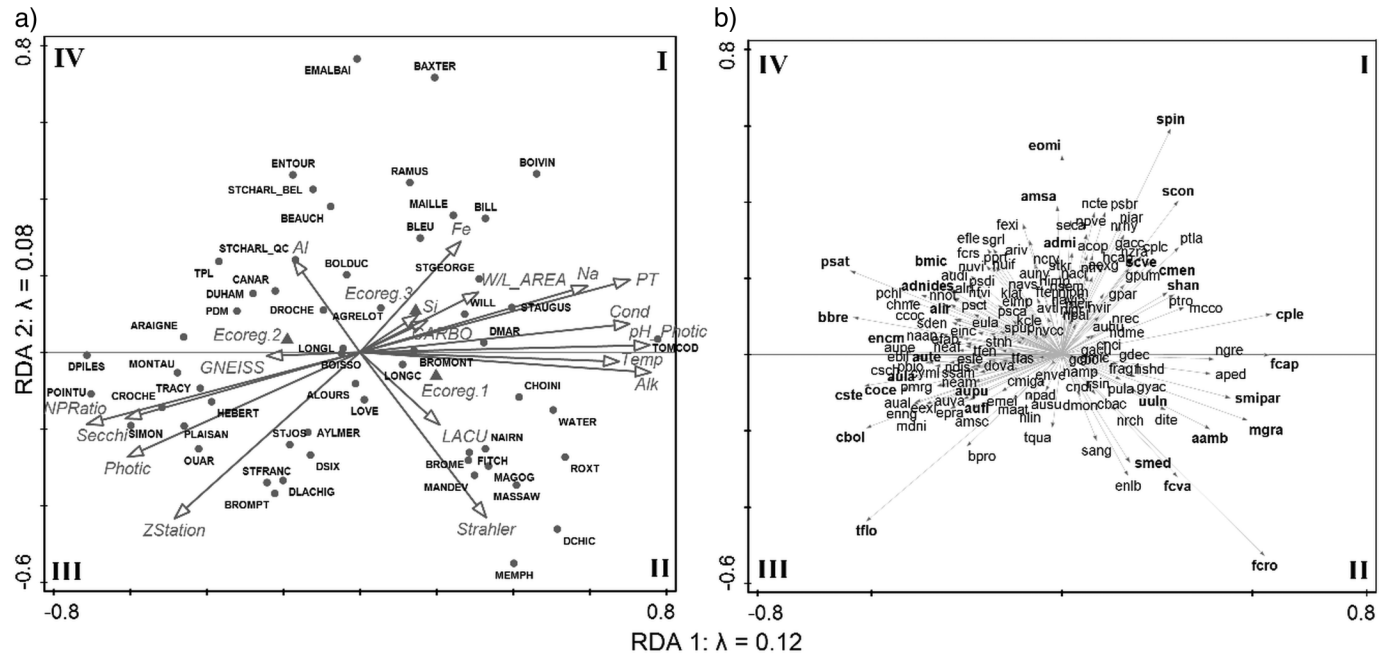
The data collected from Lac Nairne by the Réseau de suivi volontaire des lacs (RSVL) of the Ministère du développement durable, de l'environnement, de la faune et des parcs (MDDEFP, Quebec Government), along with the limnological studies of Tremblay (2005) and Labrecque et al. (2012), documented the levels of TP measured in the water column over 5 years. These measured TP concentration values were also compared with the inferred TP concentration values. The Organisation for Economic Cooperation and Development (OECD) (1982) trophic state classification system was used to compare the results of the three methods (inferred, measured, and estimated with export coefficients). The scale is divided into seven classes in terms of TP, Chla, and water transparency. This classification is based solely on TP.

## Results and discussion

### Correlations among variables

Several of the explanatory variables were correlated to one another (see Table A4, Pearson correlation matrix). The values for TP, TSS, and TiN:TP were strongly correlated. The specific conductivity and alkalinity values were also strongly correlated with each other ( $r = 0.93$ ), as the latter includes the ionic charges of anions and cations present in water, i.e.,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{+}$ ,  $\text{Na}^{+}$ ,  $\text{K}^{+}$ ,  $\text{HCO}_3^{-}$ ,  $\text{SO}_4^{2-}$ , and  $\text{Cl}^{-}$ , with alkalinity corresponding to the quantity of  $\text{CaCO}_3$  and Mg present in the water column. Linear correlations exist between specific conductivity (Cond.), sodium (Na), and alkalinity (Alk.). Furthermore, the pH values in the photic zone (pH) varied in accordance with several variables, including TP, Cond., and Alk. It is well known that the pH of the lake water has an important influence on the distribution of diatom assemblages (Smol and Stoermer 2010). It is also known that the pH in the photic zone indicates, among other things, the type of biological activity (Pourriot and Meybeck 1995). Thus, a major photosynthetic biomass generates a medium for the consumption of carbon dioxide

Fig. 3. Transformation-based canonical redundancy analysis (tb-RDA) representing the positions of the 56 lakes and 75 diatom taxa in relation to the 19 retained environmental variables.



(CO<sub>2</sub>) and the production of oxygen (O<sub>2</sub>). This effect is much less important in unproductive lakes. In our study, the pH measured in the photic zone is both a response variable (or effect variable) and an explanatory variable (or causative variable).

**Stepwise selection**

Nineteen variables were retained following stepwise selections conducted separately on the four variable types; the results are presented in Table 1. The variables that explained the most variance in the assemblages were TP concentrations, epilimnic pH values, the position of the lake in the drainage system (Strahler), and the geological regions (Ecoreg.).

**Influences of the selected variables on the diatom assemblages**

The 19 retained variables were subjected to a variance partitioning analysis to determine their unique and shared influence on diatom assemblage variability (Fig. 2). Together they explained 25.6% (adjusted R<sup>2</sup>) of the variance in the diatom assemblages of the 56 lakes (*p* value = 0.001), whereas unexplained variability (residuals) was 74.4%. In order of significance, the variables that showed the greatest influence on diatom communities were (i) causative variables TP, Al, Fe, Temp., Na, Alk., NiT:TP, Si, and Cond. (*R*<sup>2</sup><sub>adjusted</sub> = 0.048; 18.7%); (ii) geological variables Ecoreg., CARBO, GNEISS, and LACU (*R*<sup>2</sup><sub>adjusted</sub> = 0.025; 9.8%); (iii) morphometric and geographical variables Strahler, Zstation, and W/L\_Area (*R*<sup>2</sup><sub>adjusted</sub> = 0.014; 5.5%); and (iv) response variables epilimnic pH, Photic, and Secchi (*R*<sup>2</sup><sub>adjusted</sub> = 0.011; 4.3%). Shared influences represented 57.8% of the explained variance (*R*<sup>2</sup><sub>adjusted</sub> = 0.148). This shared variance can be due to, for example, correlations between the location (LAT, ALT) and the physical conditions of the water, such as temperature. Aluminum and iron concentrations tended to decrease in large lakes due to a dilution effect caused by greater water volume and depth. It is also known that the surface area of the water body has an impact on climatic and physical variables, which in turn controls biotic communities (Smol and Stoermer 2010). Other potential explanatory variables have not been considered in this study. For example, lake fetch and the force of winds both affect the mixing of the water column and internal currents (Hausmann and Pienitz 2007; Hausmann et al. 2011).

Table 2. Linear correlation coefficients obtained between the 19 environmental variables having a significant influence on the diatom assemblages and axes 1, 2, and 3 of the canonical redundancy analysis (RDA).

Environmental variable	Correlation coefficients					
	With 648 taxa			With 164 taxa (occurrence ≥ 4)		
	RDA1	RDA2	RDA3	RDA1	RDA2	RDA3
Al	0.20	-0.37	<b>0.42</b>	0.21	-0.39	<b>-0.42</b>
Fe	-0.25	-0.43	<b>0.65</b>	-0.24	-0.44	<b>-0.65</b>
Na	<b>-0.60</b>	-0.23	-0.25	<b>-0.60</b>	-0.24	0.25
Si	-0.13	-0.02	<b>0.37</b>	-0.14	-0.01	<b>-0.37</b>
TP	<b>-0.77</b>	-0.34	0.09	<b>-0.76</b>	-0.36	-0.09
Secchi	<b>0.66</b>	0.33	-0.30	<b>0.66</b>	0.34	0.30
NiT:TP	<b>0.77</b>	0.30	0.04	<b>0.77</b>	0.31	-0.04
Alk.	<b>-0.78</b>	0.05	-0.28	<b>-0.79</b>	0.04	0.28
Temp.	<b>-0.77</b>	-0.06	0.24	<b>-0.76</b>	-0.08	-0.24
pH	<b>-0.79</b>	-0.08	-0.32	<b>-0.80</b>	-0.07	0.32
Cond.	<b>-0.72</b>	-0.11	-0.31	<b>-0.72</b>	-0.12	0.31
Photic	<b>0.65</b>	0.45	-0.26	<b>0.65</b>	0.47	0.26
ZStation	0.50	<b>0.62</b>	-0.23	0.50	<b>0.62</b>	0.24
W/L_Area	-0.31	<b>-0.19</b>	0.03	-0.31	<b>-0.21</b>	-0.03
Strahler	-0.42	<b>0.58</b>	0.15	-0.42	<b>0.56</b>	-0.15
GNEISS	0.28	0.08	<b>0.36</b>	0.28	0.09	<b>-0.36</b>
CARBO	-0.18	-0.14	<b>-0.35</b>	-0.18	-0.14	<b>0.35</b>
LACU	-0.22	0.21	-0.13	-0.22	0.20	0.13
Ecoreg.	<b>0.35</b>	-0.23	-0.22	<b>0.35</b>	-0.22	0.22

Note: See Table 1 for Environmental variable definitions.

A final tb-RDA (Fig. 3) was computed using the 19 retained environmental variables to illustrate the relationship between the significant explanatory variables and their influence on the diatom assemblages. The R<sup>2</sup> of the global model indicates what proportion of the variability in the diatom assemblages is explained by the model and each canonical axis has its equivalent with an eigenvalue (λ). The R<sup>2</sup> of the global model is 0.53 and the first two axes explained together 20% of the variance (λ axis 1 = 0.12 and λ axis 2 = 0.08). Five canonical axes were significant (*p* value < 0.001). Table 2

**Table 3.** Performance statistics for the best performing WA-PLS and classic WA model for the reconstruction of total phosphorus (TP).

Total phosphorus (TP)					
WA-PLS (two-component)			Classic WA		
No. of lakes	55		No. of lakes	55	
No. of taxa (1 occurrence)	486		No. of taxa (4 occurrences)	163	
Gradient	0.9–170.6 $\mu\text{g}\cdot\text{L}^{-1}$		Gradient	0.9–170.6 $\mu\text{g}\cdot\text{L}^{-1}$	
	log10	$\mu\text{g}\cdot\text{L}^{-1}$		log10	$\mu\text{g}\cdot\text{L}^{-1}$
<b>WAPLS (two-component)</b>			<b>WAPLS (two-component)</b>		
$R^2_{\text{apparent}}$	0.9681	—	$R^2_{\text{apparent}}$	0.9128	—
Mean error	0.0057	1.01	Mean error	0.0077	1.02
RMSE	0.1379	1.37	RMSE	0.1267	1.34
$R^2_{\text{boot}}^*$	0.7111	—	$R^2_{\text{boot}}^*$	0.7349	—
Mean error	0.0104	1.02	Mean error	0.0124	1.03
Max. error	0.7861	6.11	Max. error	0.6208	4.18
RMSEP	0.3027	2.01	RMSEP	0.2774	1.89
<b>Classic WA</b>			<b>Classic WA</b>		
$R^2_{\text{apparent}}$	0.8725	—	$R^2_{\text{apparent}}$	0.8204	—
Mean error	0.0000	1.00	Mean error	0.0000	1.00
RMSE	0.1849	1.53	RMSE	0.2805	1.91
$R^2_{\text{boot}}^*$	0.6750	—	$R^2_{\text{boot}}^*$	0.7110	—
Mean error	0.0000	1.00	Mean error	0.0077	1.02
Max. error	0.6634	4.61	Max. error	0.4304	2.69
RMSEP	0.2997	1.99	RMSEP	0.2744	1.88

\*Validated by bootstrap, 1000 cycles.

shows the correlation coefficients of the first three RDA axes computed using 648 (all taxa) and 164 taxa (occurrence  $\geq 4$ ); the results are comparable.

The variables most significantly correlated with the first axis were the chemical and physical variables (Na, TP, Secchi, NiT:TP, Alk., Temp., pH, Cond., Photic) and the geological regions (Ecoreg.). The variables most correlated with the second axis were the morphometric characteristics (ZStation, W/L\_Area) and the Strahler classification (Strahler). The directions of the vectors of explanatory variables and relations between them are consistent with our knowledge of the study area and local limnology. Lakes located in the Appalachians and Saint Lawrence Lowlands are richest in nutrients, more productive and alkaline than the acidic and deeper lakes of the Canadian Shield. An increase in nutrients is also accompanied by an increase in specific conductivity and pH. The shallow lakes generally classified in the highest Strahler level, as well as those located at lower elevation and latitude, which usually are enriched in nutrients and show higher epilimnion temperatures. The lakes with a smaller surface area and larger watersheds normally yielded increased concentrations of Fe, Al, Na, and TP.

Figure 3a displays the lakes according to species, while Fig. 3b displays the species as a function of the 19 environmental variables (for clarity only 75 taxa with the best-fit are shown). The fit is measured by the squared correlation of the diatom species with the canonical axes. The codes associated with each of the taxa are listed in Table A5.

The most productive lakes are found in quadrants I and II. Quadrant I groups lakes with the smallest surface areas, which are consequently the shallowest and most enriched in nutrients and ions. The flora of these water bodies, which is often subject to daily mixing, is dominated by tycho planktonic taxa such as *Staurosirella pinnata* (spin), *Staurosira construens* (scon), *Pseudostaurosira brevistriata* (psbr), and *Staurosira construens* var. *venter* (scve). The median values of alkalinity are highest in these two quadrants. In addition, these productive and alkaline lakes are often inhabited by small species that cohabit with a significant amount of cyanobacteria blooms (field observations). Among these species, *Stephanodiscus hantzschii* (shan) and *Cyclotella meneghiniana* (cmen)

are of particularly small size. *Fragilaria capucina* (fcap) is also a dominant species in these environments.

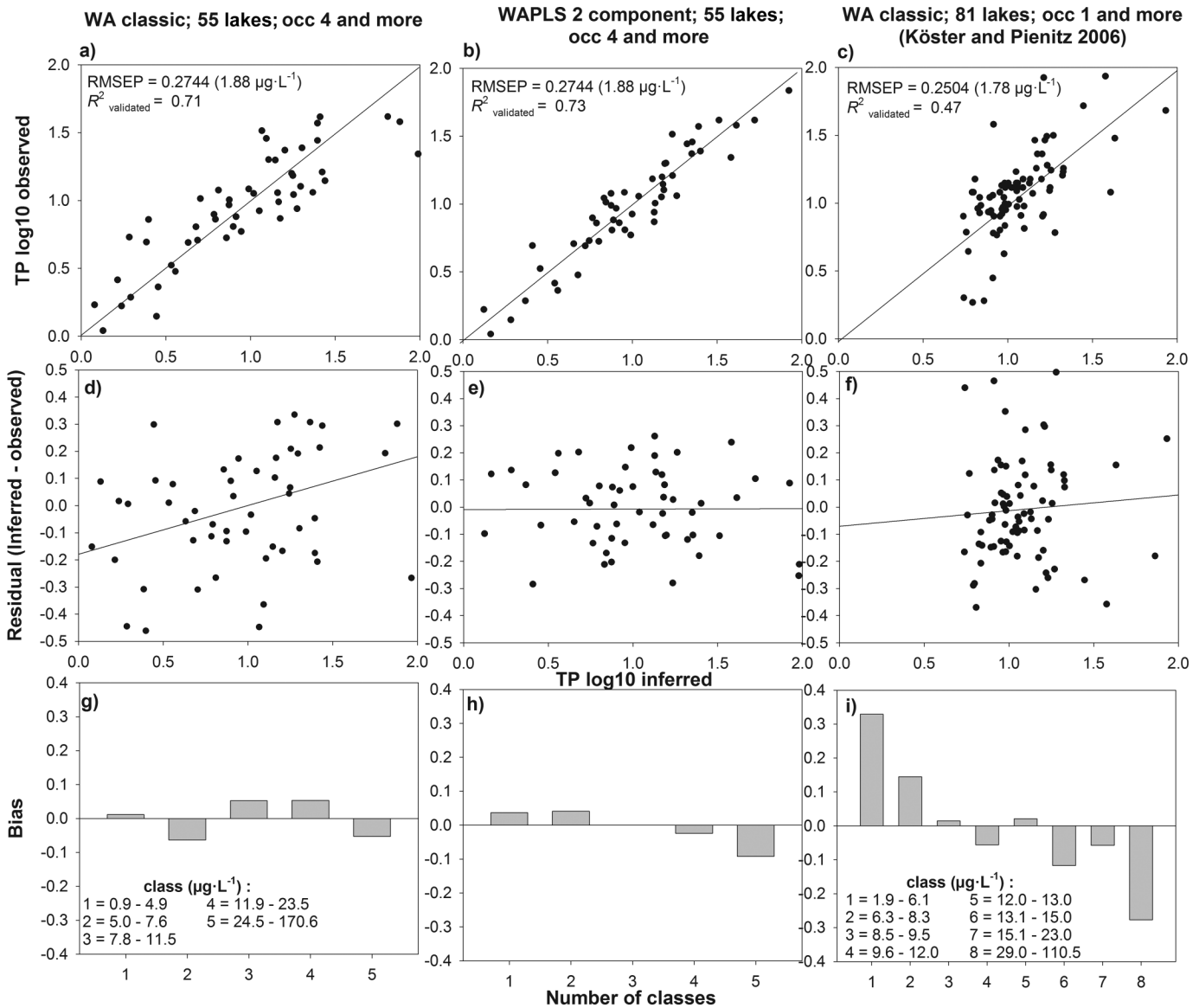
The lakes in quadrant II include mesotrophic lakes and reservoirs with large areas (lakes Memphrémagog, Massawippi, and Magog, and Reservoir Choinière), as well as small lakes that are rich in phosphorus and have high specific conductivity values (e.g., lakes Mandeville, Roxton, Waterloo, Desmarais, and des Chicots). The surface sediments of the large mesotrophic, alkaline, and dimictic lakes (Memphrémagog, Baie Fitch, Massawippi, Magog, Brome, and Nairne) are dominated by *Fragilaria crotonensis* (fcro). In the smallest lakes, *Aulacoseira ambigua* (aamb) is an important planktonic taxon. It is also interesting to note that, despite the considerable size of some of these lakes, their photic zones may reach temperatures as high as those recorded in the more shallow lakes of quadrant I.

Several halotolerant species or planktonic species that are indifferent to specific conductivity are present in quadrants I and II (Siver 1999; van Dam et al. 1994). The main species are *Fragilaria crotonensis* (fcro), *Cyclotella meneghiniana* (cmen), *Stephanodiscus hantzschii* (shan), *Stephanodiscus parvus* and *S. minutulus* (smipar), *Stephanodiscus medius* (smed), *Aulacoseira granulata* (mgra), *Cocconeis placentula* var. *euglypta* (cple), *Fragilaria capucina* var. *vaucheriae* (fcva), and *Ulnaria ulna* (uuln).

Quadrant III represents the deepest and least productive lakes. The median value of the TiN:TP ratio is much higher than that of the other quadrants because TP concentration is lower. The temperature of the photic zone and the alkalinity values are lower. The lakes of quadrant III are dominated by planktonic species that include *Cyclotella bodanica* (cbol), *Cyclotella stelligera* (cste), *Cyclotella ocellata* (coce), *Aulacoseira tenella* (aute), *Aulacoseira perglabra* var. *florimiae* (aufl), and *Tabellaria flocculosa* (tflo) (Fig. 3b). The dominant species in quadrant III are typical of low alkalinity environments (Camburn and Charles 2000).

Quadrant IV comprises lakes with low TP concentrations (e.g., Duhamel = 1.4  $\mu\text{g}\cdot\text{L}^{-1}$  and Entouré = 7.6  $\mu\text{g}\cdot\text{L}^{-1}$ ) as well as high TP concentrations (e.g., Canards = 28.6  $\mu\text{g}\cdot\text{L}^{-1}$  and Saint-Charles de Bellechasse = 173.3  $\mu\text{g}\cdot\text{L}^{-1}$ ). Despite high TP concentrations in Saint-Charles de Bellechasse due to agricultural activities, *Aulacoseira tenella* (aute) is a dominant planktonic species (15%). The latter

Fig. 4. Total phosphorus (TP)  $\log_{10}$  observed against TP  $\log_{10}$  inferred values (a, b, and c), residuals against predicted values (d, e, and f), and mean bias per interval gradient (class) (g, h, and i) for our two TP models; classic WA, two-component WA-PLS, and classic WA of Köster and Pienitz (2006). The number of classes is proportional to the number of observations (lakes/10).



is normally associated with oligotrophic conditions and low conductivities, and we suspect that the surrounding bogs explain its abundance. Furthermore, the median values for Fe and Al are higher than those of sites in quadrant III. However, the lakes of quadrant IV possess high ion values. Among these lakes, Lac Duhamel and Lac Saint-Charles de Québec rank the highest with respect to their concentrations of  $\text{Cl}^-$  and  $\text{Na}^+$ . Lake Duhamel is oligotrophic but also alkaline ( $41 \text{ mg}\cdot\text{L}^{-1} \text{ CaCO}_3$  or  $820 \mu\text{eg}\cdot\text{L}^{-1}$ ).

Overall, the RDA and variance partitioning analyses revealed that a large part ( $R^2_{\text{adjust}} = 0.048$ ; 18.7%) of the variance in the diatom assemblages is explained by the unique influence of physical and chemical variables. Inference models were developed for TP, because this variable explained, as expected, the greater independent portion (7%) of the variance in the diatom assemblages.

#### Total phosphorus (TP) inference models

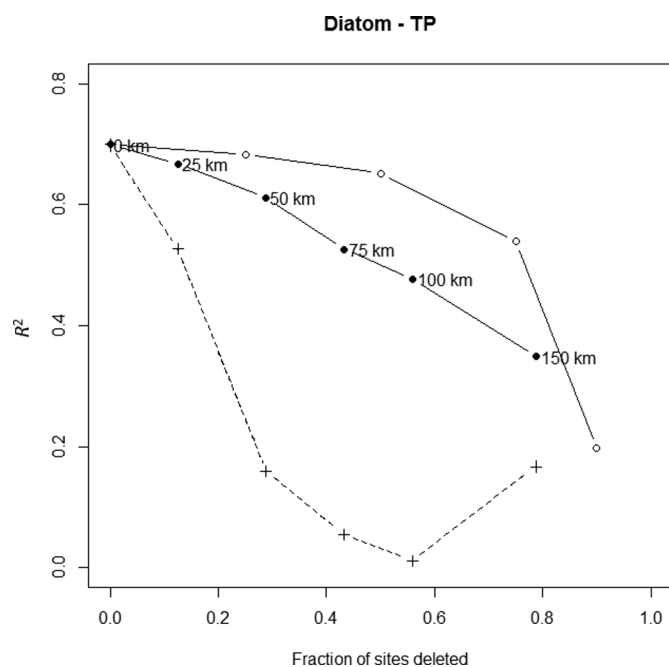
Table 3 displays the results of the optimum inference models obtained using WA with classical deshrinking and WA-PLS with the TP gradient. The number of taxa conserved as a function of the

number of occurrences and the number of retained lakes are also shown.

Lac Saint-Charles de Bellechasse was eliminated from the data set prior to the development of inference models for the reasons mentioned above. The two-component WA-PLS and the classic WA were the most efficient models (Table 3). Statistical performance of the two-component WA-PLS was better than that of the WA. However, a bias exists within the WA-PLS due to the use of inverse regression deshrinking (ter Braak and Juggins 1993). Indeed, the apparent coefficient of determination ( $R^2_{\text{apparent}}$ ) for the two-component WA-PLS (occurrences  $\geq 4$ ) is 0.91 with a  $R^2_{\text{bootstrap}}$  validation of 0.73. The root of the mean squared prediction error (RMSEP) is  $1.88 \mu\text{g}\cdot\text{L}^{-1}$ , with a maximum error of  $4.18 \mu\text{g}\cdot\text{L}^{-1}$  and a mean error of  $1.03 \mu\text{g}\cdot\text{L}^{-1}$ . Furthermore, the coefficient of determination  $R^2_{\text{apparent}}$  of the classic WA (occurrences  $\geq 4$ ) is 0.82 and the  $R^2_{\text{bootstrap}}$  is 0.71. The RMSEP is  $1.88 \mu\text{g}\cdot\text{L}^{-1}$ , with a maximum error of  $2.69 \mu\text{g}\cdot\text{L}^{-1}$  and a mean error of  $1.02 \mu\text{g}\cdot\text{L}^{-1}$ . These models cover a gradient from 0.9 to  $170.6 \mu\text{g}\cdot\text{L}^{-1}$ , representing



**Fig. 5.** The effect on transfer function  $R^2$  of deleting sites at random (mean of 10 trials; open circles); from the geographical neighbourhood of the test site (closed circles); or that are environmentally most similar (dotted line with crosses) during cross-validation.



ultra-oligotrophic ( $<4 \mu\text{g}\cdot\text{L}^{-1}$ ) to hyper-eutrophic ( $>100 \mu\text{g}\cdot\text{L}^{-1}$ ) conditions (OECD 1982). Thus, the RMSEP of the two models equals 1% of the TP gradient. The inference model based on 163 species (occurrences  $\geq 4$ ) rather than 486 species (occurrences  $\geq 1$ ) performs slightly better. In addition, its maximum error was smaller.

Figure 4 illustrates the performance statistics for our two TP models with occurrence  $\geq 4$  and classic WA of Köster and Pienitz (2006). The number of intervals is proportional to the number of observations (lakes/10). The bias of TP classes provides a better assessment of the model performance than the regression of residuals. According to the validation of classic WA, the model tends to slightly overestimate or to underestimate the TP values (bias  $< 1.3 \mu\text{g}\cdot\text{L}^{-1}$  or  $[0.1]$  TP  $\log_{10}$ ) depending on class. Based on the two-component WA-PLS, the biases are of the same order of magnitude as those of the classic WA model. The TP optimum values for individual species estimated from the classic WA model and the regression coefficients  $r$  species parameters (or species parameters “optima”) from the WA-PLS model are listed in Table S1. Finally, the model by Köster and Pienitz (2006) generates biases that are greater at the gradient extremes than in our models (bias  $< 2.0 \mu\text{g}\cdot\text{L}^{-1}$  or  $[0.3]$  TP  $\log_{10}$ ).

#### Does spatial autocorrelation affect the performance of our inference models in an inappropriate way?

As proposed by Telford and Birks (2009), we calculated the effect of deleting sites from a geographic neighbourhood on the  $R^2$  to gauge the effect of spatial autocorrelation on the inference model (WA). Figure 5 presents these results. The assumption is that if there was spatial autocorrelation, deleting analog sites before applying the cross-validation test would cause a drop of  $R^2$  with this test. Several cross-validation tests were produced for each geographic distance (25–150 km). The line with open circles shows the effect of deleting sites at random (mean of 10 trials), whereas the line with closed circles shows the effect of deleting sites from the geographical neighbourhood of the test-site on the transfer function  $R^2$ . The dotted line with crosses shows the effect of deleting sites that are environmentally most similar on the transfer function  $R^2$ .

Thus, spatial autocorrelation had no effect since  $R^2$  did not drop in a drastic way but decreased gradually with a greater fraction of sites deleted. These results are consistent with Telford and Birks’ (2009) diatom-pH model. As they mentioned, spatial autocorrelation may be a much greater challenge for a data set involving an open system like an ocean than the sum of several closed systems, such as our lake systems. Based on our results, spatial autocorrelation does not seem to play an important role in our inference TP model.

#### TP reconstruction based on fossil diatom assemblages from Lac Nairne

The TP values inferred for the sediment sequence from Lac Nairne (Roy, 2012) were obtained using the previously calibrated classic WA and WA-PLS (55 lakes and occurrences  $\geq 4$ ), to reconstruct its trophic evolution over the past 2000 years. First, an analog test between the modern diatom assemblages of the model data sets and the fossil data of the Lac Nairne sediment sequence was conducted (Fig. A1). The inferred TP values were compared with those measured in the water column for the summer periods of 2004, 2005, 2008, 2009, and 2010, as well as with TP values estimated using export coefficients of the watershed’s phosphorus flowing towards the lake (Labrecque 2013). Table 4 displays the data set, as well as the TP values inferred from the diatom-based models.

The general trends of the data obtained from the different methods concerning the chronological sequence led to the same conclusions. This means that the TP values, either inferred or estimated using TP export coefficients, showed the same result with respect to the trophic status evolution of the lake through time. The bias associated with different classes of TP shows an inverse trend between the classic WA model (Fig. 4g) and the two-component WA-PLS model (Fig. 4h). Consequently, using the average of the two models resulted in better inferred values, as these were closer to those obtained using the TP export coefficients, except for the base of the core. The bedrock of the Lac Nairne watershed is composed of 25% carbonate rocks which could explain the higher natural TP content than normally observed in Canadian Shield lakes. The presence of carbonate rocks also likely explains that Lac Nairne waters are alkaline ( $43 \text{ mg}\cdot\text{L}^{-1}$  of  $\text{CaCO}_3$  or  $860 \mu\text{eq}\cdot\text{L}^{-1}$ ). TP export coefficients developed by the Lakeshore Capacity Model (Paterson et al. 2006) have been calibrated based on south-central Ontario lakes that are also located on the Canadian Shield, yet alkaline lakes like Lac Nairne were not present except for one. The particular geological setting of Lac Nairne could explain export coefficients that underestimate natural TP concentrations in the lake. Bennion et al. (2005) compared the natural phosphorus concentrations of 62 lakes from Britain with the inference model based on the diatom assemblages and TP export coefficients model, and in 74% of cases the concentration obtained using diatoms was higher than that obtained with the export coefficients model.

The evolution of trophic state in Lac Nairne reveals that prior to any human activities in the catchment (50 cm depth in sediment record), the concentration of TP varied between  $5.3$  and  $8.1 \mu\text{g}\cdot\text{L}^{-1}$  (Table 4). The inferred data classify Lac Nairne as oligotrophic during this period. By the 1930s, anthropogenic enrichment had already begun. Between 1940 and 1958 (6.8–5.2 cm depth), TP concentrations inferred by models and estimated with the TP export coefficients attained their maximum, ranging between  $21.1$  and  $26.0 \mu\text{g}\cdot\text{L}^{-1}$  ( $>20 \mu\text{g}\cdot\text{L}^{-1}$ ), which rates the lake as meso- to eutrophic. In the 1980s, both the inferred and estimated TP concentrations varied between  $17.0$  and  $18.1 \mu\text{g}\cdot\text{L}^{-1}$ . These lower TP values represent a mesotrophic state ( $<20 \mu\text{g}\cdot\text{L}^{-1}$ ). From 2004 to 2010, the inferred and estimated TP values ranged between  $17.4$  and  $20 \mu\text{g}\cdot\text{L}^{-1}$ . The limnological study of Lac Nairne conducted in 2010 effectively classifies the lake as mesotrophic, confirming the overall similarity of the results. The inferred TP values, as well as

**Table 4.** Total phosphorus (TP) values measured, estimated, and inferred for the water column of Lac Nairne, Charlevoix.

TP measured in water column							TP inferred (occurrence $\geq 4$ )						TP estimated		
Sampling year	TP values ( $\mu\text{g}\cdot\text{L}^{-1}$ )						TP values ( $\mu\text{g}\cdot\text{L}^{-1} \pm 1.9$ )						With exportation coefficient for phosphorus ( $\mu\text{g}\cdot\text{L}^{-1}$ )		
	Mean	June	July	August	September	October	Sediment trap			Two-component WA-PLS	Classic WA	Average			
2010*	16.6	—	8.1	7.3	34.3	—							14 (20) <sup>¶</sup>		
2010 <sup>†</sup>	15.2	Every two weeks in the summer (n=10)													
2009 <sup>‡</sup>	8.7	9.8	9.3	7.0	—	—	Sediment trap			10.9	12.6	11.8			
							Sample location in sediment core N-01								
							Core depth (cm)								
							Pb <sup>210</sup> dating								
							Upper			Lower	Middle				
2009 <sup>‡</sup>	8.7	9.8	9.3	7.0	—	—	0	0.2	2009	2007	2008	15.8	19.0	17.4	
2005 <sup>‡</sup>	12.0	—	12.0	—	—	—	—	—	—	—	—	—	—	—	
2005 <sup>§</sup>	19.8	—	17.0	5.0	36.0	21.0	—	—	—	—	—	—	—	—	
2004 <sup>‡</sup>	8.9	—	6.5	11.3	—	—	0.4	0.6	2004	2003	2004	16.0	19.3	17.6	
1980	Aerial photographs to document the land						3.4	3.6	1981	1979	1980	16.4	19.8	18.1	17
							4	4.2	1974	1971	1973	17.1	21.8	19.4	
							4.4	4.6	1969	1966	1967	19.7	24.0	21.8	
1964	Aerial photographs to document the land						—	—	—	—	—	—	—	—	23
							5	5.2	1960	1957	1958	22.0	28.5	25.2	
							5.4	5.6	1954	1952	1953	21.4	27.3	24.3	
1950	Aerial photographs to document the land						—	—	—	—	—	—	—	—	26
							6.0	6.2	1947	1945	1946	19.2	24.3	21.7	
							6.4	6.6	1943	1941	1942	22.1	27.3	24.7	
							6.6	6.8	1941	1939	1940	18.8	23.5	21.1	
							7.0	7.2	1937	1936	1937	17.2	20.7	18.9	
							7.4	7.6	1935	1933	1934	21.1	24.8	23.0	
							8.0	8.2	1930	1929	1930	17.4	20.2	18.8	
							8.4	8.6	1927	1926	1927	14.4	17.3	15.9	
							9.0	9.2	1920	1922	1921	14.8	17.7	16.2	
							C <sup>14</sup> dating								
Before anthropogenic influence	Woodlands (natural conditions)						49.5	50.0	-900 AD			7.4	8.9	8.1	5.3

\*This study.

<sup>†</sup>Labrecque et al. (2012).<sup>‡</sup>Réseau de suivi volontaire des lacs (RSVL) of Ministère du Développement Durable, de l'Environnement et des Parcs (MDDEP 2012).<sup>§</sup>Tremblay (2005).<sup>¶</sup>Included release phosphorus in water by sediment.

**Table 5.** Performance statistics for the WA and the two-component WA-PLS models developed using the data set of Köster and Pienitz (2006), including 188 taxa and 81 lakes.

	WA	Two-component WA-PLS
$R^2_{\text{apparent}}$	0.71	0.80
$R^2_{\text{boot}}$ *	0.47	0.54
Mean error	1.02 $\mu\text{g}\cdot\text{L}^{-1}$	1.01 $\mu\text{g}\cdot\text{L}^{-1}$
Max. error	3.28 $\mu\text{g}\cdot\text{L}^{-1}$	3.53 $\mu\text{g}\cdot\text{L}^{-1}$
RMSE	1.58 $\mu\text{g}\cdot\text{L}^{-1}$	1.38 $\mu\text{g}\cdot\text{L}^{-1}$
RMSEP	1.78 $\mu\text{g}\cdot\text{L}^{-1}$	1.74 $\mu\text{g}\cdot\text{L}^{-1}$

\*Validated by bootstrap, 1000 cycles.

those estimated using the export coefficients, indicate a recovery of Lac Nairne that was favored by decreasing agricultural activities in the watershed in the 1970s, along with a natural and gradual reforestation process that occurred over the same time period. However, it is important to note that the concentrations measured in the water column between 2004 and 2010 were highly variable, as sampling intervals and areas were not constant from year to year. Thus, an analysis of the fossil diatom assemblages for a given period (~1 year) may provide more precise and relevant data concerning the overall trophic state of the lake than instrumental measurements if the latter were not taken in a systematic manner.

#### Comparison of the reconstruction models using different data sets

To determine the performance of the new models in this study and to compare the values inferred by another model, the historical TP levels of Lac Nairne were also reconstructed using the data sets of Köster and Pienitz (2006). First, an analogy test between the modeled diatom assemblages of Lac Nairne was conducted (Fig. A2). The fossil assemblages of Lac Nairne were below the 95th percentile dissimilarity threshold of the data sets. However, three important species that were found at various levels in the Lac Nairne sediment core record were absent from the data set, including *Cyclotella bodanica* var. *lemanica*, *Fragilaria capucina*, and *Stephanodiscus medius*. Thus, the lower levels of the Lac Nairne sediment sequence (31–59 cm core depths), which include relative abundances of *Cyclotella bodanica* var. *lemanica* reaching up to 20%, were not included in the model reconstruction.

The TP model of Köster and Pienitz (2006) is based on diatom assemblages sampled from the surface sediments of 81 lakes in New England, selected from a data set of more than 200 lakes from the northeastern United States (Dixit et al. 1999). The TP gradient of Köster and Pienitz (2006) varies between 1.9 and 110.5  $\mu\text{g}\cdot\text{L}^{-1}$ , as compared with the gradient of our model (0.9 to 170.6  $\mu\text{g}\cdot\text{L}^{-1}$ ). Thirty taxa from the diatom assemblages of Lac Nairne were common in both data sets.

The performance statistics of the inference models developed using the classic WA and two-component WA-PLS along with the data set of Köster and Pienitz (2006) are presented in Table 5, whereas the classic WA model performance is shown in Figs. 4c, 4f, and 4i. The results of the reconstruction are illustrated in Fig. 6. Values inferred by this model for the upper levels were generally lower than those inferred using our data set. These results are consistent with the model bias (Fig. 4i, class 4). The classic WA developed using the data set of Köster and Pienitz (2006) yielded values closer to those inferred by the two-component WA-PLS of the present study. Overall, the profiles show the same trends over time. The divergent values can be attributed to the greater dissimilarity between Lac Nairne assemblages and those of Köster and Pienitz (2006).

The average bias illustrated in Fig. 4i shows that inferred values tended to underestimate TP concentrations for values  $\geq 9.5 \mu\text{g}\cdot\text{L}^{-1}$ . The number of taxa used to infer TP values is smaller in the model of Köster and Pienitz (2006) than in our model. This could explain

why the most important peaks are associated with specific species, such as *Stephanodiscus hantzschii*, *Staurosirella pinnata*, *Aulacoseira granulata*, *Eolimna minima*, *Cocconeis placentula* var. *euglypta*, and *Planolithidium lanceolatum*. In summary, the TP values inferred with this study's data set for the upper sediment levels are of the same order of magnitude as those measured by Labrecque et al. (2012) and comparable to the estimates produced using the export coefficients related to the human occupation of the Lac Nairne watershed (Labrecque 2013).

#### Comparison between the TP values inferred and the semiquantitative index

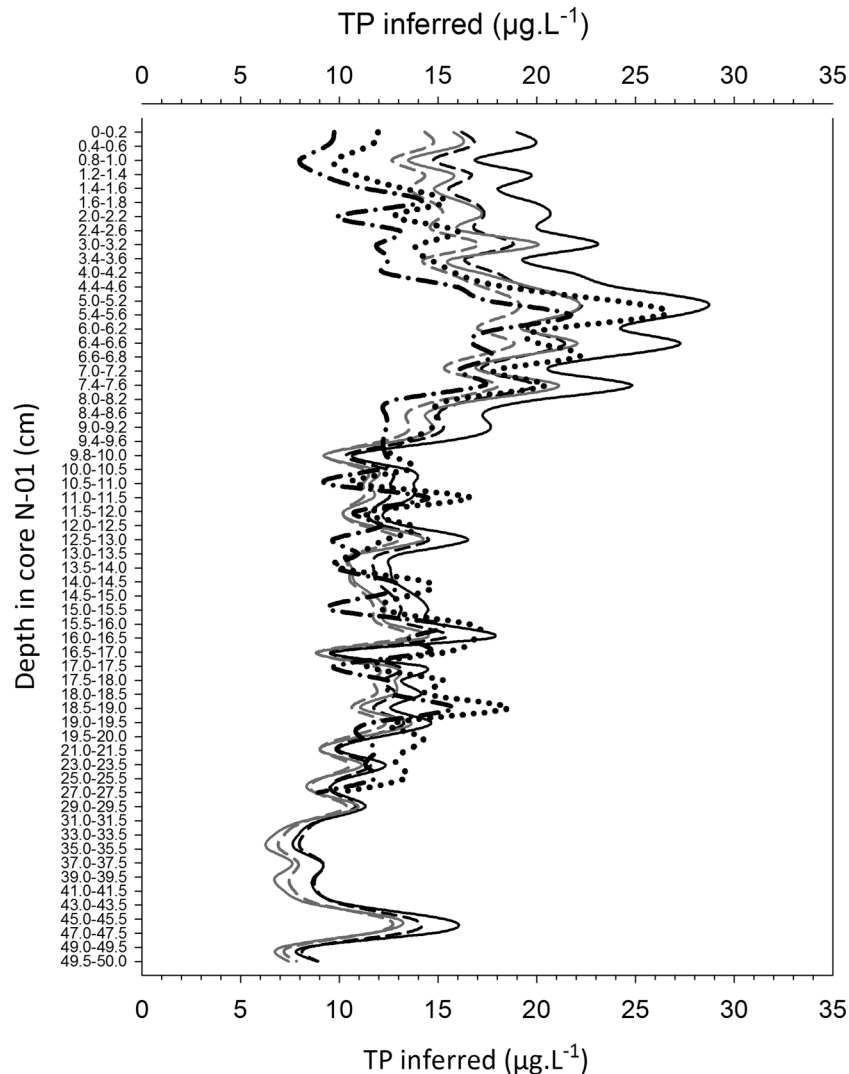
Figure 7 shows the TP values inferred from this study's models, as well as those developed using the index calculated with the fossil diatoms identified and listed by Roy (2012); all values are thus based on the diatom assemblages of Lac Nairne. A Pearson correlation test was conducted between the three sections of the profile, based on clusters from Ward's minimum variance method (Ward 1963), with the fossil diatom assemblages of Roy (2012). Section 1 comprised depths from 0 to 8.2 cm, section 2 from 8.4 to 29.5 cm, and section 3 from 31 to 50 cm. The similarity between the profiles was good except for the uppermost zone (zone 1) where the trend between 4.4 cm to the core top was almost reverse ( $r = 0.14$ ;  $p$  value = 0.545). The correlation ( $r$ ) between the curves of zone 2 was 0.63 ( $p$  value < 0.000) and the correlation ( $r$ ) between curves in zone 3 was 0.89 ( $p$  value < 0.000). The low correlation between the curves of zone 1 is explained by the extreme influence of *Fragilaria crotonensis* and *Aulacoseira ambigua* in the semiquantitative index and the underestimation of the *Stephanodiscus* spp. signal, associated with eutrophic conditions. In contrast, the TP inference models better predicted the overall increase of *Fragilaria crotonensis* in zone 1, as they reduce the importance of this bloom-forming taxon. This effect was also observed in the inverted peak from 4.4 to 4.6 cm depth. At this level, *Fragilaria crotonensis* drastically decreased in the assemblages and consequently reduced values of the semiquantitative index.

#### Choice of inference model

The choice of an inference model is first based on its statistical performance. In addition, it should be subjected to a validation test. The WA-PLS excels for finding a linear solution that minimizes the sum of squares errors when calibrating a model. For example, if a null-type inference model is constructed using a diatom assemblage from a modern data set with the TP values being randomly mixed, the results show that the null-type WA-PLS will perform well notwithstanding, as the  $R^2$  of the model is 0.91 and the RMSE is 1.70  $\mu\text{g}\cdot\text{L}^{-1}$ . In contrast, when the  $R^2$  is validated using a bootstrap (1000 cycles) or leave-one-out cross-validation test, the results indicate that the model has no predictive value ( $R^2_{\text{validated}} = 0.004$ ). Figure A3 shows a graphical representation of the results of the two-component null-type WA-PLS.

Second, model accuracy must be defined as the ability to quantitatively infer historical changes in TP measurements (Hall et al. 1997). The results of the study also show that the inference models with the best statistical performance are not necessarily the best for reconstructing a variable for a given lake. Third, the potential inappropriate effect of spatial autocorrelation must be assessed. Fourth, the inferred values must be tested against instrumental observations to ensure the adequacy of the model variable. Our modelling tests with sodium (Na) showed that it was correlated with TP, and that it also explained a portion of the variability in the assemblages, independently of other variables, and that model performance was acceptable. On the other hand, the reconstructions of the top-bottom (modern-fossil) assemblages in 56 cores have shown that Na was a surrogate of TP. Such situations occur when the wrong inferred value systematically adopts the behaviour of the variable that has real influence. Finally, as suggested by Racca and Prairie (2004), only plots of residuals against

**Fig. 6.** Total phosphorus (TP) profiles inferred using the Lac Nairne diatom assemblages and the data sets of this study (long-dashed black line, classic WA and occurrences  $\geq 1$ ; solid black line, classic WA and occurrences  $\geq 4$ ; long-dashed grey line, two-component WA-PLS and occurrences  $\geq 1$ ; solid grey line, two-component WA-PLS and occurrences  $\geq 4$ ) and those of Köster and Pienitz (2006) (dotted black line, classic WA; dash-dot black line, two-component WA-PLS).



predicted values are useful to evaluate the general inherent trends of numerical models. In addition, the bias by class is most appropriate for discriminating the best model for a specific lake.

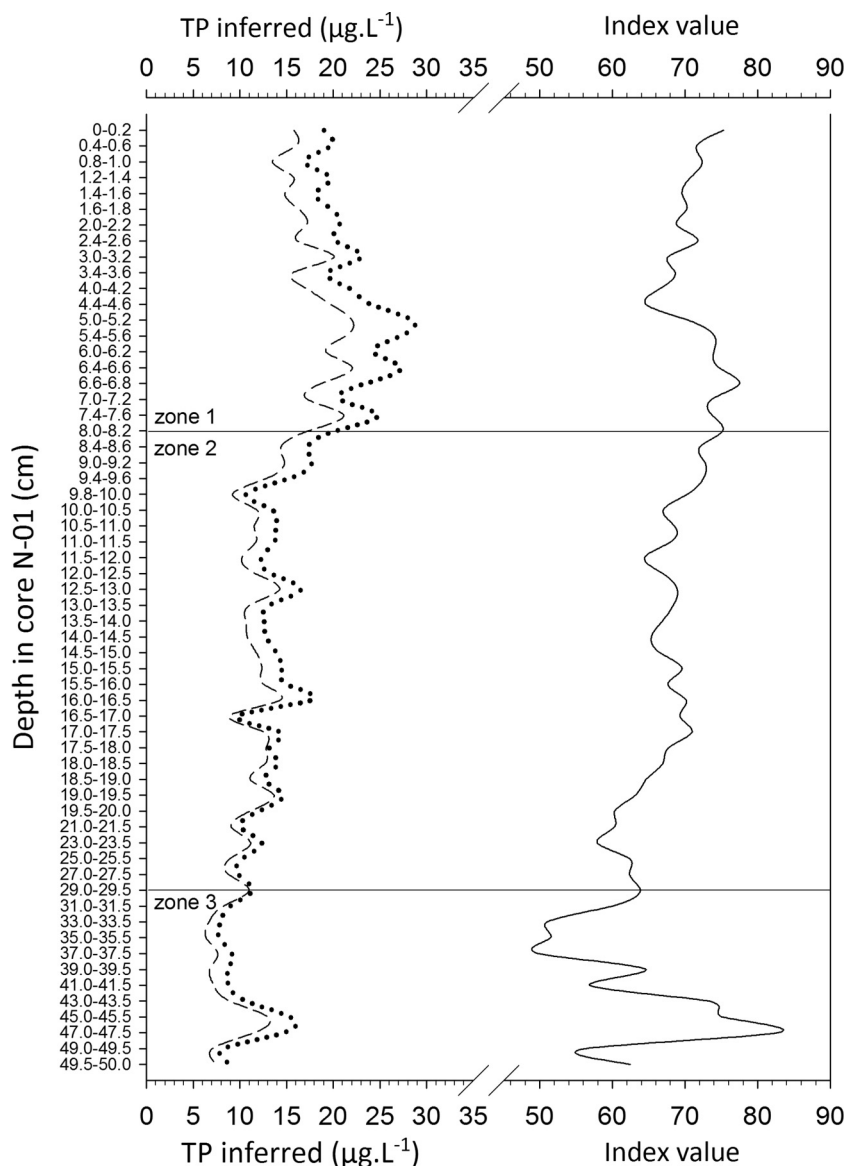
#### Reliability of inference model

The growing limnological knowledge of the Lac Nairne ecosystem indicates that the TP inference model provides a coherent and accurate depiction of this evolution. In addition, a comparison of the inferred TP levels with those obtained using the semi-quantitative index emphasizes that inference models incorporate complex interactions between communities in different trophic states by using weightings applied to the taxa of an assemblage based on a set of observations. For example, *Fragilaria crotonensis* possesses relative abundance values within an assemblage that are much higher than those of *Stephanodiscus hantzschii*, despite the environmental conditions being optimal for each taxon. This phenomenon occurs because *Fragilaria crotonensis* produces blooms and grows in conditions that are less restrictive than *Stephanodiscus hantzschii*, while the latter is subjected to strong pressures from other planktonic communities in eutrophic to hyper-eutrophic environments. Figure A4 shows the number of *Fragilaria crotonensis* and *Stephanodiscus hantzschii* counted in relation to the phos-

phorus levels measured in the 56 lakes of the data set. The figure shows that, despite the optimal conditions, *Stephanodiscus hantzschii* did not attain the high levels of abundance that *Fragilaria crotonensis* did. The inference model seems to better weigh the variability between the different growth rates between taxa and different trophic states.

A large proportion of the variance in the diatom assemblages of each data set is likely explained by ecological information that is still unknown. In contrast, the results of the comparison tests in this study indicate that the inference models have the capacity to accurately reconstruct trends. However, the precision of the inference models' output will largely depend on the level of precision of the data used to create the models. The modelling of complex systems necessarily involves some degree of simplification. The main factor that contributes to imprecisions in the models is the study object itself. The responses of a living organism to a given variable can be highly variable, even in unicellular organisms. A second factor concerns the precision of the measured variables, which are the averages of three point measurements taken over a specific time frame (summer, spring, or annual). Thus, by accepting these potential factors of imprecision in modelling efforts, it is

**Fig. 7.** Total phosphorus (TP) profiles of the Lac Nairne water column inferred using the study's models with occurrences  $\geq 4$  (dotted line, classic WA; long-dashed line, two-component WA-PLS) and a semiquantitative diatom index (solid line).



**Table 6.** Statistical performance of total phosphorus (TP) inference models from various regions in the world.

Gradient ( $\mu\text{g}\cdot\text{L}^{-1}$ )	Model type and season	$R^2_{\text{apparent}}$ and $R^2_{\text{boot}}$ *	RMSEP ( $\mu\text{g}\cdot\text{L}^{-1}$ )	Location and no. of lakes (n)	Source
1-171	WA, summer	0.82, 0.71	1.88	Southern Quebec (55)	This study
30-550	WA, annual	—, 0.75	1.32	Yangtze River, China (43)	Yang et al. 2008
2-171	WA total, annual	—, 0.50	1.58	New Zealand (53)	Reid 2005
4-54	Two-component WA-PLS, spring	0.69, 0.47	1.55	Southeastern Ontario, Canada (101)	Werner and Smol 2005
3-52	Two-component WA-PLS, summer	0.89, 0.51	3.20	Abitibi and Mauricie, Quebec, Canada (76)	Philibert and Prairie 2002
24-1145	Two-component WA-PLS, annual	0.86, 0.37	1.91	Denmark (29)	Bradshaw et al. 2002
7-370	WA, annual	0.75, 0.47	1.74	Sweden (43)	Bradshaw and Anderson 2001
8-8740	WA, summer	—, 0.55	2.20	Northeastern United States (238)	Dixit et al. 1999
3-83	WA, summer	0.77, 0.52	1.70	Alaska, United States (51)	Gregory-Eaves et al. 1999

\*Validated by bootstrap, 1000 cycles.

unreasonable to attach too much importance to the precision of the absolute inferred values. In most cases, the general trends of the evolution of an environmental variable provide insightful and sufficient information, as we have shown via the relationship between

the trophic classes and TP concentrations of Lac Nairne. The relevance of the models relies upon their capacity to provide important information as a function of time. Thus, the challenge is to ensure that the data set is adequate for each fossil assemblage. In this way,

the methodological imprecision (prediction error) will be constant for all sediment core levels and variations and hence be comparable. Likewise, the accuracy is high enough to use these inference models to determine natural TP concentrations in lakes and thereby estimate their support capacity (or limit capacity) to prevent trophic changes (CCME 2004).

### Conclusion

The results shown here demonstrate that it is possible to attain the objectives defined at the start of this study. The performance of the TP transfer functions is comparable to other TP inference models described in the literature (Table 6). Our model yields the best validated  $R^2$ , except for the  $R^2$  of Yang et al. (2008), which is based on a data set with 86 species and an RMSEP that is comparable to the other models.

The results of this study indicate that the new TP models should be useful tools to help guide managers toward concerted action. By documenting natural background conditions and the degree to which they have changed, the models provide important information regarding the sustainable management of a lake or group of lakes. In addition, they allow for realistic recovery goals concerning TP concentrations in the water column to be set.

The Lac Nairne scenario shows that it is impossible to document the trophic evolution of a lake using only a few point summer measurements of TP in the water column, even when conducted over several years (7 years). In contrast, fossil diatoms provide a reliable signal of change in TP over time. Thus, due to the decrease in agriculture and natural reforestation of the watershed since the 1960s, Lac Nairne is currently in a less eutrophic state than it was in 1950, despite the growth of resorts around the lake.

This paper has shown that a paleolimnological approach can provide information about the relationships that exist between certain activities in a watershed and the deterioration of water quality and destruction of habitat over time. In certain cases, this approach is the only available option to investigate past impacts, in addition to being the most direct and efficient method. This approach is, therefore, appropriate for the documentation of the evolution of natural conditions in southern Quebec lakes, with which the support capacity of each lake can be determined.

Overall, it is important to remember that a model can be statistically efficient without being a realistic model, which may be explained by the model capacity itself (e.g., over-fitting) or by the structure of its data (e.g., spatial autocorrelation). Beyond the technical validation of inference models, ecological knowledge and logistic constraints often force us to work with sets of incomplete and imperfect data. It seems impossible to study all the variables of influence as it is an inherent property of biological communities to respond to a whole range of environmental conditions. At the same time, it seems impossible to use predictor variables that are not influenced by others. Thus, model tests are ultimately the best way to assess their validity and to define their limits. In addition, the comparison of TP values of the two independent training data sets used in our study confirmed that diatoms reliably infer TP concentrations and not a surrogate.

Regarding a more technical aspect, our results support points raised by Birks (1995), namely that the average of inferred values obtained with two different methods often yielded the best result. Finally, it seems that violating the “rule of thumb” as expressed by ter Braak (1988) does not prevent the development of interesting TP models. Would it not be appropriate to reassess this rule? If the lakes would have been more carefully selected to minimize variability between other variables of influence, then our model would not have been very useful in practice, given that the TP gradient would have been much shorter.

### Acknowledgements

This project results from a collaboration between Université Laval and CIMA+ and benefitted from funding provided by CIMA+,

the Natural Sciences and Engineering Research Council (NSERC) of Canada (Discovery grant of R.P.), and logistical support from Centre d'études Nordiques (CEN). Furthermore, we are grateful to the municipalities in the study region and the residents of the lake catchments. Likewise, we would like to thank the management of the Patrimoine Écologique et des parcs, MDDEFP, particularly Daniel Blais, for his collecting of data concerning the hydrological basins of 56 lakes in southern Quebec, and the management team responsible for the monitoring of the quality of the environment from MDDEFP, particularly from the Réseau du Suivi Volontaire des Lacs (RSVL). Finally, we also thank Paul Jasinski for translating the manuscript, and Dermot Antoniadis and two anonymous referees for their constructive comments.

### References

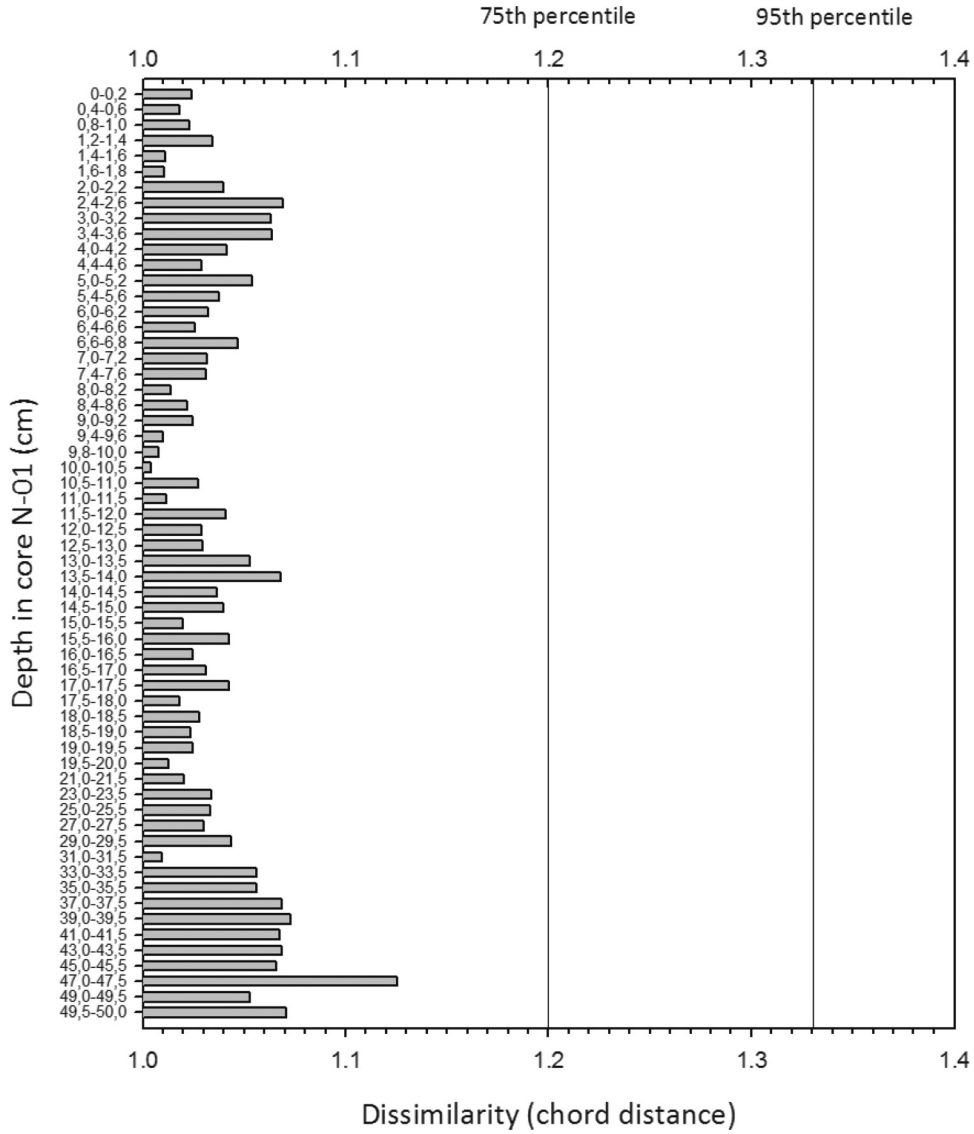
- Adler, S. 2010. PALTRAN: WA, WA-PLS and MW for paleolimnology. R package version 1.3-0. Available from <http://www.biologie.uni-rostock.de/abt/botanik/AG-Phykologie/index-engl.htm>.
- Bennion, H., Johnes, P., Ferrier, R., Phillips, G., and Haworth, E. 2005. A comparison of diatom phosphorus transfer functions and export coefficient models as tools for reconstructing lake nutrient histories. *Freshw. Biol.* **50**: 1651–1670. doi:10.1111/j.1365-2427.2005.01428.x.
- Birks, H.J.B. 1995. Quantitative palaeoenvironmental reconstructions. In *Statistical modelling of quaternary science data*. Technical Guide 5. Edited by D. Maddy and J.S. Brew. Quaternary Research Association, Cambridge. pp. 161–254.
- Birks, H.J.B. 1998. Numerical tools in palaeolimnology — progress, potentialities, and problems. *J. Paleolimnol.* **20**(4): 307–332. doi:10.1023/A:1008038808690.
- Birks, H.J.B., Lotter, A.F., Juggins, S., and Smol, J.P. 2012. Tracking environmental change using lake sediments. Vol. 5: Data handling and numerical techniques. Springer Dordrecht, Heidelberg, New York, London. 745 p.
- Blanchet, F.G., Legendre, P., and Borcard, D. 2008. Forward selection of explanatory variables. *Ecology* **89**: 2623–2632. PMID:18831183.
- Borcard, D., Legendre, P., and Drapeau, P. 1992. Partialling out the spatial component of ecological variation. *Ecology*, **73**: 1045–1055. doi:10.2307/1940179.
- Bradshaw, E.G., and Anderson, N.J. 2001. Validation of a diatom-phosphorus calibration set for Sweden. *Freshw. Biol.* **46**: 1035–1048. doi:10.1046/j.1365-2427.2001.00732.x.
- Bradshaw, E.G., Anderson, N.J., Jensen, J.P., and Jeppesen, E. 2002. Phosphorus dynamics in Danish lakes and the implications for diatom ecology and palaeoecology. *Freshw. Biol.* **47**: 1963–1975. doi:10.1046/j.1365-2427.2002.00938.x.
- Camburn, K.E., and Charles, D.F. 2000. Diatoms of low-alkalinity lakes in the northeastern United States. The Academy of Natural Sciences of Philadelphia. 152 p.
- Canadian Council of Ministers of the Environment (CCME). 2004. Recommandations canadiennes pour la protection de la vie aquatique : le phosphore : cadre canadien d'orientation pour la gestion des réseaux hydrographiques. Dans : Recommandations canadiennes pour la qualité de l'environnement, 2004, CCME, Winnipeg. 7 p.
- Centre d'expertise en analyse environnementale du Québec (CEAEQ). 2008. Détermination du phosphore total dans les eaux naturelles : minéralisation au persulfate; méthode colorimétrique automatisée; procédures adaptées pour le phosphore en teneur élevée et à l'état de trace, MA. 303 – P 5.0, Rév. 3. 29 p.
- Dixit, S.S., Smol, J.P., Charles, D.F., Hugues, R.M., Paulsen, S.G., and Collins, G.B. 1999. Assessing water quality changes in the lakes of the northeastern United States using sediment diatoms. *Can. J. Fish. Aquat. Sci.* **56**(1): 131–152. doi:10.1139/f98-148.
- Fallu, M.-A., Allaire, N., and Pienitz, R. 2000. Freshwater diatoms from northern Québec and Labrador (Canada). *Bibliotheca Diatomologica* Vol. 45. J. Cramer, Berlin/Stuttgart. 200 p.
- Gregory-Eaves, I., Smol, J.P., Finney, B.P., and Edwards, M.E. 1999. Diatom-based transfer functions for inferring past climatic and environmental changes in Alaska, U.S.A. *Arctic Antarctic Alpine Res.* **31**(4): 353–365.
- Hall, R.I., and Smol, J.P. 1992. A weighted-averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes. *Freshw. Biol.* **27**: 417–434. doi:10.1111/j.1365-2427.1992.tb00551.x.
- Hall, R.I., Leavitt, P.R., Smol, J.P., and Zirnelt, N. 1997. Comparison of diatoms, fossil pigments and historical records as measures of lake eutrophication. *Freshw. Biol.* **38**: 401–417. doi:10.1046/j.1365-2427.1997.00251.x.
- Hausmann, S., and Pienitz, R. 2007. Seasonal climate inferences from high-resolution modern diatom data along a climate gradient: a case study. *J. Paleolimnol.* **38**: 73–96. doi:10.1007/s10933-006-9061-2.
- Hausmann, S., Larocque-Tobler, I., Richard, P.J.H., Pienitz, R., St-Onge, G., and Fye, F. 2011. Diatom-inferred wind activity at Lac du Sommet, southern Québec, Canada: a multiproxy paleoclimate reconstruction based on diatoms, chironomids and pollen for the past 9500 years. *Holocene*, **21**: 925–938. doi:10.1177/0959683611400199.

- Hollister, J., and Milstead, W.B. 2010. Using GIS to estimate lake volume from limited data. *Lake Reservoir Manage.* **26**: 194–199. doi:10.1080/07438141.2010.504321.
- Horne, A.J., and Goldman, C.R. 1994. *Limnology*. 2nd ed. McGraw Hill Inc., New York. 576 p.
- Idso, S.B., and Gilbert, R.G. 1974. On the universality of the Poole and Atkins Secchi disk-light extinction equation. *J. Appl. Ecol.* **11**: 399–401. doi:10.2307/2402029.
- Juggins, S. 2013a. Quantitative reconstructions in palaeolimnology: new paradigm or sick science? *Quat. Sci. Rev.* **64**: 20–32. doi:10.1016/j.quascirev.2012.12.014.
- Juggins, S. 2013b. rioja: analysis of quaternary science data. R package version 0.5-6. Available from <http://www.staff.ncl.ac.uk/staff/stephen/juggins/>.
- Juggins, S., and Birks, H.J.B. 2012. Quantitative environmental reconstruction from biological data. In *Tracking environmental change using lake sediments*. Vol. 5: Data handling and numerical techniques. Edited by H.J.B. Birks, A.F. Lotter, S. Juggins, and J.P. Smol. Springer Dordrecht, Heidelberg, New York, London. pp. 431–494.
- Kalff, J. 2002. *Limnology: inland water ecosystems*. Prentice Hall, Upper Saddle River, N.J. 608 p.
- Koenings, J.P., and Edmundson, J.A. 1991. Secchi disk and photometer estimates of light regimes in Alaskan lakes: effects of yellow color and turbidity. *Limnol. Oceanogr.* **36**: 91–105. doi:10.4319/lo.1991.36.1.0091.
- Köster, D., and Pienitz, R. 2006. Late-Holocene environmental history of two New England ponds: natural dynamics versus human impacts. *Holocene*, **16**(4): 519–532. doi:10.1191/0959683606h1947rp.
- Krammer, K., and Lange-Bertalot, H. 1986. *Bacillariophyceae*, 1. Teil: Naviculaceae. Gustav Fischer Verlag, 876 p.
- Krammer, K., and Lange-Bertalot, H. 1988. *Bacillariophyceae*, 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. Gustav Fischer Verlag, 596 p.
- Krammer, K., and Lange-Bertalot, H. 1991a. *Bacillariophyceae*, 3. Teil: Centrales, Fragilariaceae, Eunotiaceae. Gustav Fischer Verlag, 576 p.
- Krammer, K., and Lange-Bertalot, H. 1991b. *Bacillariophyceae*, 4. Teil: Achnantheaceae. Gustav Fischer Verlag, 437 p.
- Labrecque, V. 2013. Quand le passé refait surface : caractérisation des apports en phosphore au lac Nairne, Charlevoix, Québec. M.Sc. thesis, Department of Geography at Université Laval, Québec. 107 p.
- Labrecque, V., Nürnberg, G., Tremblay, R., and Pienitz, R. 2012. Caractérisation de la charge interne de phosphore du lac Nairne, Charlevoix (Québec). *Rev. Sci. Eau*, **25**(1): 77–92. doi:10.7202/1008537ar.
- Lavoie, I., Hamilton, P.B., Campeau, S., Grenier, M., and Dillon, P.J. 2008. Guide d'identification des diatomées des rivières de l'est du Canada. Presse de l'Université du Québec. 241 p.
- Legendre, P., and Birks, H.J.B. 2012. From classical to canonical ordination. In *Tracking environmental change using lake sediments*. Vol. 5: Data handling and numerical techniques. Edited by H.J.B. Birks, A.F. Lotter, S. Juggins, and J.P. Smol. Springer Dordrecht, Heidelberg, New York, London. pp. 201–248.
- Legendre, P., and De Cáceres, M. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecol. Lett.* **16**: 951–963. doi:10.1111/ele.12141. PMID:23809147.
- Legendre, P., and Gallagher, E. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**: 271–280. doi:10.1007/s004420100716.
- Legendre, P., and Legendre, L. 2012. *Numerical ecology*. 3rd English ed. Elsevier Science BV, Amsterdam.
- Li, T., and Ducruc, J.P. 1999. Les provinces naturelles. Niveau I du cadre écologique de référence du Québec. Ministère de l'Environnement. 90 p.
- Mason, C.F. 1991. *Biology of freshwater pollution*. 2nd ed. Longman Scientific and Technical, Longman Group UK Ltd. 351 p.
- Ministère du Développement Durable, de l'Environnement et des Parcs (MDDEP). 2012. Available from <http://www.mddep.gouv.qc.ca/eau/rsvl/methodes.htm> [accessed 13 November 2012].
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, M., and Wagner, H. 2013. VEGAN: community ecology package. R package version 2.0-8. Available from <http://cran.r-project.org/>.
- Organization for Economic Cooperation and Development (OECD). 1982. *Eutrophication of waters — monitoring, assessment, and control*. Paris, OECD Cooperative Programme on Monitoring of Inland Waters. 154 p.
- Overpeck, J.T., Webb, T., and Prentice, I.C. 1985. Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs. *Quat. Res.* **23**(1): 87–108. doi:10.1016/0033-5894(85)90074-2.
- Paterson, A.M., Dillon, P.J., Hutchinson, N.J., Futter, M.N., Clark, B.J., Mills, R.B., Reid, R.A., and Scheider, W.A. 2006. A review of the components, coefficients and technical assumptions of Ontario's lakeshore capacity model. *Lake Res. Manage.* **22**: 7–18. doi:10.1080/07438140609353880.
- Philibert, A., and Prairie, Y.T. 2002. Diatom-based transfer functions for western Quebec lakes (Abitibi and Haute Mauricie): the possible role of epilimnetic CO<sub>2</sub> concentration in influencing diatom assemblages. *J. Paleolimnol.* **27**(4): 465–480. doi:10.1023/A:1020372724266.
- Pienitz, R., Smol, J.P., and Birks, H.J.B. 1995. Assessment of freshwater diatoms as quantitative indicators of past climatic change in the Yukon and Northwest Territories, Canada. *J. Paleolimnol.* **13**: 21–49. doi:10.1007/BF00678109.
- Pourriot, R., and Meybeck, M. 1995. *Limnologie générale*. Masson, Paris, France. 956 p.
- R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org>.
- Racca, J.M.J., and Prairie, Y.T. 2004. Apparent and real bias in numerical transfer functions in palaeolimnology. *J. Paleolimnol.* **31**: 117–124. doi:10.1023/B:JOPL.0000013283.51877.72.
- Reid, M. 2005. Diatom-based models for reconstructing past water quality and productivity in New Zealand lakes. *J. Paleolimnol.* **33**(1): 13–38. doi:10.1007/s10933-004-0424-2.
- Roy, A.-J. 2012. Reconstitution de l'état trophique du lac Nairne, Charlevoix, QC. M.Sc. thesis, Department of Geography, Université Laval, Québec. 122 p.
- Schindler, D.W. 1974. Eutrophication and recovery in experimental lakes: implications for lake management. *Science*, **184**: 897–899. doi:10.1126/science.184.4139.897. PMID:17782381.
- Siver, P.A. 1999. Development of paleolimnological inference models for pH, total nitrogen and specific conductivity based on planktonic diatoms. *J. Paleolimnol.* **21**: 45–60. doi:10.1023/A:1008047906094.
- Siver, P.A., and Kling, H. 1997. Morphological observations of *Aulacoseira* using scanning electron microscopy. *Can. J. Bot.* **75**(11): 1807–1835. doi:10.1139/b97-894.
- Smol, J.P., and Stoermer, E.F. 2010. *The diatoms: applications for the environmental and earth sciences*. 2nd ed. Cambridge University Press. 667 p.
- Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. *Trans. Am. Geophys. Union*, **8**(6): 913–920.
- Telford, R.J. 2013. palaeoSig: significance tests for palaeoenvironmental reconstructions. R package version 1.1-2. Available from <http://cran.stat.sfu.ca/> [accessed 25 November 2013].
- Telford, R.J., and Birks, H.J.B. 2005. The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. *Quat. Sci. Rev.* **24**: 2173–2179. doi:10.1016/j.quascirev.2005.05.001.
- Telford, R.J., and Birks, H.J.B. 2009. Evaluation of transfer functions in spatially structured environments. *Quat. Sci. Rev.* **28**: 1309–1316. doi:10.1016/j.quascirev.2008.12.020.
- ter Braak, C.J.F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, **41**(4): 859–873. doi:10.2307/2530959.
- ter Braak, C.J.F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**(5): 1167–1179. doi:10.2307/1938672.
- ter Braak, C.J.F. 1988. CANOCO — a FORTRAN program for canonical community ordination by partial detrended canonical correspondence analysis, principal components analysis and redundancy analysis (version 2.1). Technical Report LWA-88-02. Agricultural Mathematics Group, Wageningen.
- ter Braak, C.J.F., and van Dam, H. 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia*, **178**: 209–223. doi:10.1007/BF00006028.
- ter Braak, C.J.F., and Juggins, S. 1993. Weighted averaging partial least-squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia*, **269**: 485–502.
- Tremblay, R. 2005. Bilan du suivi limnologique et recommandations pour freiner l'eutrophisation accélérée du lac Nairne et minimiser l'impact de ces symptômes pour les villégiateurs. Robert Hamelin et associés, QE841-04-038, Lévis, Que., Canada. 27 p.
- van Dam, H., Mertens, A., and Sinkeldam, J. 1994. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Neth. J. Aquat. Ecol.* **28**: 117–133. doi:10.1007/BF02334251.
- Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* **58**: 236–244. doi:10.1080/01621459.1963.10500845.
- Werner, P., and Smol, J.P. 2005. Diatom-environmental relationships and nutrient transfer functions from contrasting shallow and deep limestone lakes in Ontario, Canada. *Hydrobiologia*, **533**: 145–173. doi:10.1007/s10750-004-2409-6.
- Yang, X.D., Anderson, N.J., Dong, X.H., and Shen, J. 2008. Surface sediment diatom assemblages and epilimnetic total phosphorus in large, shallow lakes of the Yangtze floodplain: their relationships and implications for assessing long-term eutrophication. *Freshw. Biol.* **53**(7): 1273–1290. doi:10.1111/j.1365-2427.2007.01921.x.

## Appendix

Appendix tables and figures appear on the following pages.

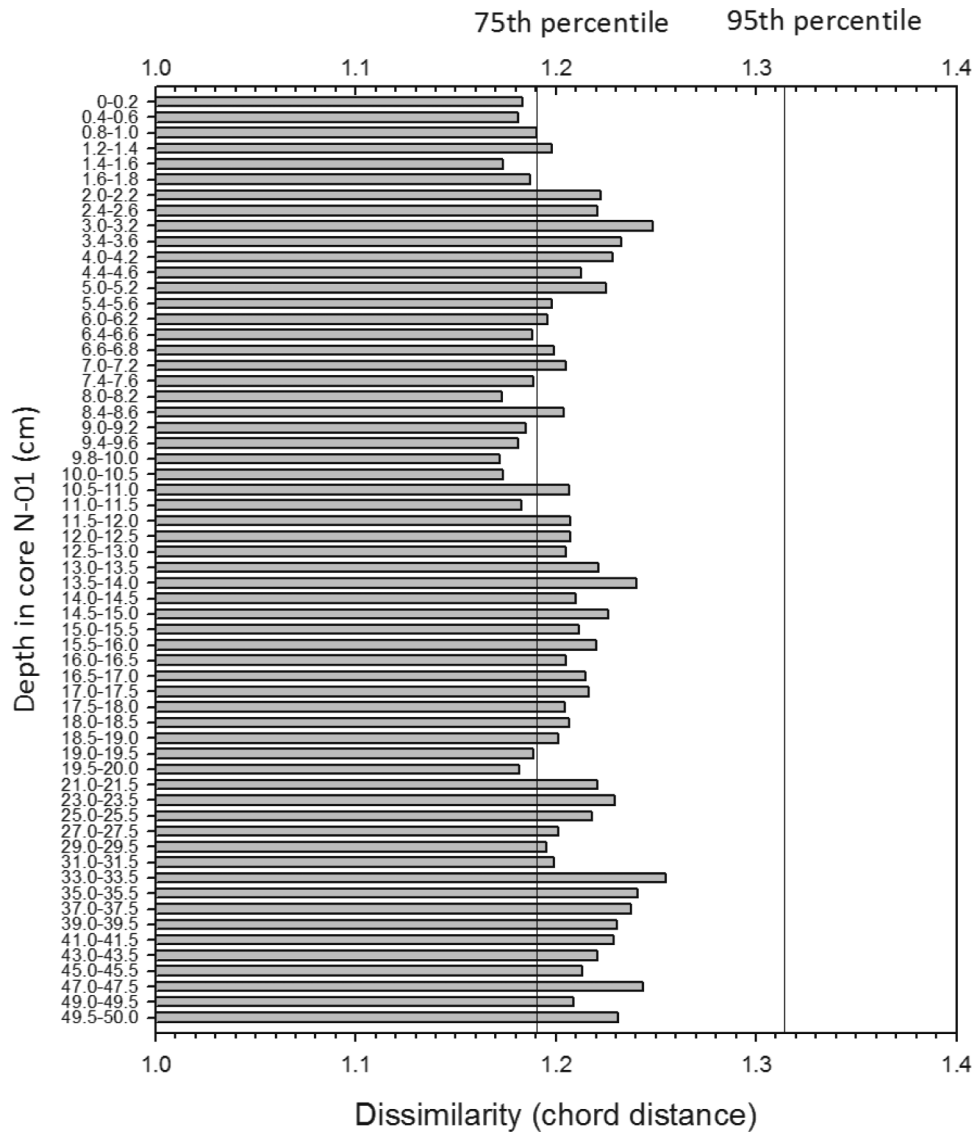
Fig. A1. Dissimilarity between the fossil assemblages of Lac Nairne and the data set assemblages of this study.



Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by Depository Services Program on 07/31/14  
For personal use only.

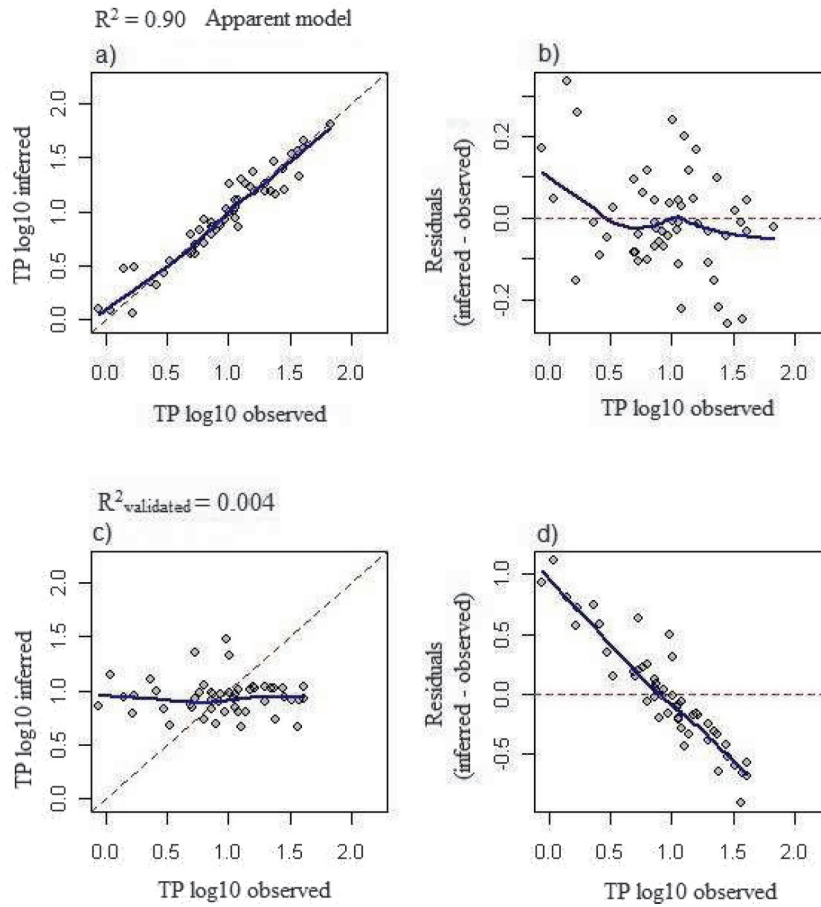


Fig. A2. Dissimilarity between the fossil assemblages of Lac Nairne and the data set assemblages of Köster and Pienitz (2006).

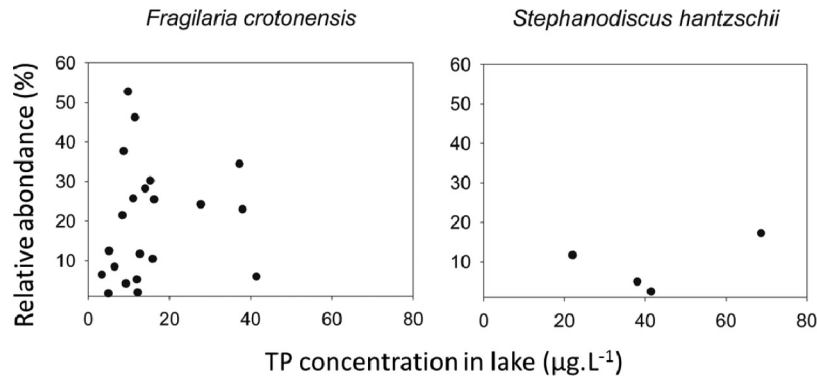


Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by Depository Services Program on 07/31/14  
For personal use only.

**Fig. A3.** Regression and residuals of the two component null-type WA-PLS created using the randomly mixed  $\log_{10}$  total phosphorus (TP) values, along with 52 lakes and 468 taxa. Panels *a* and *b* show the apparent model and *c* and *d* the validated model.



**Fig. A4.** Relative abundance (%) of *Fragilaria crotonensis* and *Stephanodiscus hantzschii* counted in all of the lakes (56) of the data set used to develop the total phosphorus (TP) inference models.



**Table A1.** Lake characteristics.

No.	Lake	Acronym	Sampling year	Latitude (°)	Longitude (°)	Altitude (m)	Geological region
1	aux Grelots	AGRELOT	2010	45.9741	-71.0013	329	A
2	à l'Ours	ALOURS	2010	45.9594	-74.0559	320	CS
3	aux Araignées	ARAIGNE	2009	45.4748	-70.8176	400	A
4	Aylmer	AYLMER	2009	45.8236	-71.3610	226	A
5	Baxter	BAXTER	2010	46.4526	-70.8366	385	A
6	Beauchamp	BEAUCH	2009	46.2660	-74.0988	541	CS
7	Bill	BILL	2010	46.5744	-72.9279	162	CS
8	Bleu	BLEU	2009	45.3602	-72.9256	65	SLL
9	Boissonneault	BOISSO	2010	45.6008	-71.9218	202	A
10	Boivin	BOIVIN	2010	45.4030	-72.6940	112	A
11	Bolduc	BOLDUC	2010	46.0148	-71.1085	310	A
12	Brome	BROME	2009	45.2480	-72.5145	195	A
13	Bromont	BROMONT	2009	45.2660	-72.6710	129	A
14	Brompton	BROMPT	2009	45.4332	-72.1449	237	A
15	Canards	CANAR	2010	46.8303	-70.8173	80	SLL
16	Choinière	CHOINI	2009	45.4279	-72.6000	140	A
17	Croche	CROCHE	2010	45.9918	-74.0051	356	CS
18	des Chicots	DCHIC	2010	46.7989	-72.5216	143	SLL
19	de l'Achigan	DLACHIG	2010	45.9289	-73.9676	200	CS
20	Desmarais	DMAR	2009	45.4586	-72.1170	271	A
21	des Piles	DPILES	2010	46.6478	-72.8004	160	CS
22	Des Roches	DROCHE	2009	46.9218	-71.2530	300	CS
23	des Six	DSIX	2010	46.5368	-72.9062	200	CS
24	Duhamel	DUHAM	2010	46.1438	-74.6415	240	CS
25	Étang Malbaie	EMALBAI	2010	47.7294	-70.7453	710	CS
26	Entouré	ENTOUR	2010	47.6835	-70.8070	698	CS
27	Baie Fitch	FITCH	2009	45.0676	-72.2226	197	A
28	Hébert	HEBERT	2010	47.6958	-71.0018	853	CS
29	Long (Charlevoix)	LONGC	2009	47.7403	-70.4174	240	CS
30	Long (Lanaudière)	LONGL	2009	46.1681	-73.6724	250	CS
31	Lovering	LOVE	2009	45.1710	-72.1564	240	A
32	Magog	MAGOG	2009	45.3054	-72.0412	183	A
33	Maillé	MAILLE	2010	45.8672	-74.0016	223	CS
34	Mandeville	MANDEV	2009	46.3716	-73.3243	142	CS
35	Massawippi	MASSAW	2009	45.2170	-72.0110	156	A
36	Memphrémagog	MEMPH	2009	45.1233	-72.2438	197	A
37	Montaubois	MONTAU	2010	45.9227	-74.0718	351	CS
38	Nairne	NAIRN	2009	47.6857	-70.3497	210	CS
39	Ouareau	OUARO	2009	46.2862	-74.1433	382	CS
40	Pied-des-Monts	PDM	2009	47.7623	-70.4124	240	CS
41	Plaisant	PLAISAN	2010	46.5009	-72.9440	175	CS
42	Pointu	POINTU	2010	47.7175	-70.7534	755	CS
43	Rat-Musqué	RAMUS	2009	47.7261	-70.3985	240	CS
44	Roxton	ROXT	2009	45.4670	-72.6535	121	A
45	Simon	SIMON	2010	46.8975	-72.0317	171	CS
46	Saint-Augustin	STAUGUS	2010	46.7507	-71.3911	40	SLL
47	Saint-Charles (Québec)	STCHARL_QC	2009	46.9346	-71.3850	146	CS
48	Saint-Charles (Bellechasse)	STCHARL_BEL	2010	46.7803	-70.9630	80	SLL
49	Saint-François	STFRANC	2009	45.9024	-71.1706	272	A
50	Saint-Georges	STGEORGE	2010	45.6448	-71.8838	244	A
51	Saint-Joseph	STJOS	2009	46.9059	-71.6370	156	CS
52	Tomcod	TOMCOD	2009	45.5377	-72.0378	202	A
53	Trois-Petits-Lacs	TPL	2009	47.0285	-71.4222	300	CS
54	Tracy	TRACY	2010	45.9266	-74.0655	347	CS
55	Waterloo	WATER	2009	45.3340	-72.5192	206	A
56	William	WILL	2009	46.1207	-71.5686	192	A

Note: CS, Southern Laurentian Mountains; A, Southern Appalachian Mountains; SLL, Saint Lawrence Lowlands.

**Table A2.** Analysis of lake environmental and morphometric characteristics.

Variable	Unity	Polymictic lakes, <i>n</i> = 21			Dimictic lakes, <i>n</i> = 35			All lakes, <i>n</i> = 56		
		Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
Cl	mg·L <sup>-1</sup>	14.81	0.32	120.33	6.15	0.19	41.74	9.35	0.19	120.33
SO <sub>4</sub>	mg·L <sup>-1</sup>	12.01	1.26	146.00	4.00	1.21	10.52	7.06	1.21	146.00
NO <sub>3</sub>	mg·L <sup>-1</sup>	0.17	0.00	0.56	0.28	0.00	1.70	0.22	0.00	1.70
COD	mg·L <sup>-1</sup>	6.84	3.31	17.53	4.42	0.76	11.43	5.33	0.76	17.53
NiI	mg·L <sup>-1</sup>	0.44	0.15	1.73	0.26	0.07	0.78	0.31	0.07	1.73
Al	mg·L <sup>-1</sup>	0.03	0.01	0.08	0.02	0.00	0.10	0.03	0.00	0.10
Ca	mg·L <sup>-1</sup>	12.70	2.82	45.87	7.74	1.01	31.75	9.42	1.01	45.87
Fe	mg·L <sup>-1</sup>	0.13	0.01	0.48	0.06	0.00	0.58	0.09	0.00	0.58
K	mg·L <sup>-1</sup>	1.11	0.23	5.47	0.58	0.10	2.45	0.74	0.10	5.47
Mg	mg·L <sup>-1</sup>	3.21	0.51	18.14	2.45	0.21	14.10	2.54	0.21	18.14
Mn	mg·L <sup>-1</sup>	0.02	0.00	0.07	0.01	0.00	0.08	0.01	0.00	0.08
Na	mg·L <sup>-1</sup>	10.47	0.96	78.81	3.96	0.47	23.00	6.42	0.47	78.81
S	mg·L <sup>-1</sup>	4.08	0.69	47.21	1.43	0.49	3.54	2.43	0.49	47.21
Si	mg·L <sup>-1</sup>	1.42	0.16	3.90	1.48	0.15	4.18	1.47	0.15	4.18
Chla	µg·L <sup>-1</sup>	29.38	0.77	95.62	4.81	0.42	23.07	13.62	0.42	95.62
TP	µg·L <sup>-1</sup>	45.96	6.40	173.33	7.97	0.87	22.03	21.91	0.87	173.33
Secchi	m	1.6	0.4	3.4	4.8	1.3	25.0	3.7	0.4	25.0
TiN:TP		20	2	145	47	9	176	35	2	145
TSS	mg·L <sup>-1</sup>	7.7	0.3	36.4	1.4	0.1	3.5	3.8	0.1	36.4
Alk.	mg·L <sup>-1</sup> CaCO <sub>3</sub>	35.5	6.8	114.7	26.1	1.8	97.0	28.5	1.8	114.7
Temp.	°C	20.48	15.68	22.98	17.55	9.17	22.79	18.51	9.17	22.98
pH		8.15	6.69	9.50	7.59	6.37	9.39	7.75	6.37	9.50
Cond.	mS·cm <sup>-1</sup>	0.15	0.03	0.71	0.09	0.01	0.26	0.11	0.01	0.71
DO	mg·L <sup>-1</sup>	7.72	5.97	12.51	7.26	4.81	10.24	7.43	4.81	12.51
Photic	m	2.5	0.7	5.0	7.9	2.2	30.0	6.0	0.7	30.0
Thermo	m	3.5	1.0	14.0	6.9	3.0	15.0	5.5	1.0	15.0
ZStation	m	4.4	0.8	14.0	18.3	6.0	30.0	12.9	0.8	30.0
Volume	m <sup>3</sup>	3 837 064	34 085	31 827 608	110 182 651	135 293	1 611 834 210	73 936 410	34 085	1 611 834 210
Lake_area	Km <sup>-2</sup>	1.32	0.03	8.69	7.50	0.03	67.13	5.25	0.03	67.13
V/S	m	1.85	0.11	6.06	9.37	1.51	38.05	6.68	0.11	38.05
W_area	Km <sup>-2</sup>	40.6	0.9	259.0	203.9	0.2	1739.6	139.7	0.2	1739.6
W/L_Area	Km <sup>-2</sup>	52	3	377	26	3	130	35	3	377
Qs	M·year <sup>-1</sup>	31	2	211	18	2	90	22	2	211
Diatom_SED	Valves·g <sup>-1</sup> dry sediment	3.40E+06	2.26E+05	1.38E+07	3.74E+06	4.47E+05	1.43E+07	3.70E+06	2.26E+05	1.43E+07
Diatom_MO	Valves·g <sup>-1</sup> organic matter	1.48E+07	1.88E+06	4.06E+07	1.52E+07	1.43E+06	5.71E+07	1.54E+07	1.43E+06	5.71E+07

**Table A3.** Results of Shapiro–Wilk normality tests on the raw and transformed data.

Environmental variable	Raw data		Transformation	Transformed data	
	W	p		W	p
Chlorine (Cl)	0.443	0.00	log 10	0.953	0.03
Sulfate (SO <sub>4</sub> )	0.220	0.00	log 10	0.812	0.00
Nitrate (NO <sub>3</sub> )	0.614	0.00	x <sup>0.5</sup>	0.895	0.00
Dissolved organic carbon (DOC)	0.878	0.00	x <sup>0.5</sup>	0.966	0.11
Total dissolved inorganic nitrogen (TiN)	0.587	0.00	log 10	0.954	0.04
Aluminium (Al)	0.841	0.00	x <sup>0.5</sup>	0.945	0.03
Calcium (Ca)	0.732	0.00	log 10	0.983	0.60
Iron (Fe)	0.734	0.00	log 10	0.969	0.15
Potassium (K)	0.619	0.00	log 10	0.982	0.55
Magnesium (Mg)	0.657	0.00	log 10	0.988	0.87
Manganese (Mn)	0.580	0.00	log 10	0.972	0.21
Sodium (Na)	0.433	0.00	log 10	0.972	0.17
Sulfur (S)	0.226	0.00	log 10	0.794	0.00
Silica (Si)	0.941	0.01	x <sup>0.5</sup>	0.983	0.61
Chlorophyll <i>a</i> (Chl <sub>a</sub> )	0.607	0.00	log 10	0.969	0.15
Total phosphorus (TP)	0.509	0.00	log 10	0.977	0.34
TiN:TP ratio (NPRatio)	0.842	0.00	x <sup>0.5</sup>	0.955	0.04
Total suspended solids (TSS)	0.500	0.00	log 10	0.988	0.85
Alkalinity (Alk.)	0.866	0.00	x <sup>0.5</sup>	0.970	0.17
Turbidity (Secchi)	0.646	0.00	log 10	0.980	0.45
Temperature (Temp.)	0.945	0.01	—	—	—
Hydrogen potential (pH)	0.979	0.44	—	—	—
Conductivity (Cond.)	0.632	0.00	log 10	0.988	0.85
Dissolved oxygen (DO)	0.940	0.01	x <sup>0.5</sup>	0.964	0.09
Depth photic zone (Photic)	0.762	0.00	log 10	0.987	0.78
Depth thermocline (Thermo)	0.940	0.01	log 10	0.960	0.06
Maximal depth of bassin (ZStation)	0.912	0.00	x <sup>0.5</sup>	0.952	0.03
Lake volume (Volume)	0.332	0.00	log 10	0.952	0.21
Lake area (Lake_area)	0.468	0.00	log 10	0.983	0.59
Volume/Lake_Area (V/S)	0.777	0.00	x <sup>0.5</sup>	0.956	0.04
Watershed area (W_area)	0.502	0.00	log 10	0.973	0.25
Watershed area / lake area (W/L_Area)	0.486	0.00	log 10	0.972	0.21
Water load per unit area (Qs)	0.548	0.00	log 10	0.983	0.61
No. of diatom valves per gram of dry sediment (Diatom_SED)	0.804	0.00	log 10	0.985	0.69
No. of diatom valves per gram of organic matter (Diatom_MO)	0.845	0.00	log 10	0.983	0.61

Note: W critic = 0947 if  $\alpha = 0.05$ ; W critic = 0930 if  $\alpha = 0.01$ ;  $n = 56$ .

**Table A4.** Pearson correlation matrix computed on the 57 explanatory variables selected.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Causative variables														
	Cl	SO <sub>4</sub>	NO <sub>3</sub>	COD	NiT	Al	Ca	Fe	K	Mg	Mn	Na	S	Si	TP
Cl	1.00														
SO <sub>4</sub>	0.64	1.00													
NO <sub>3</sub>	0.10	0.19	1.00												
COD	0.18	0.08	0.31	1.00											
NiT	0.31	0.43	0.47	0.68	1.00										
Al	-0.30	-0.23	0.20	0.52	0.34	1.00									
Ca	0.78	0.69	0.18	0.25	0.46	-0.31	1.00								
Fe	-0.01	-0.23	0.09	0.64	0.41	0.61	-0.03	1.00							
K	0.73	0.67	0.28	0.35	0.63	-0.19	0.83	0.12	1.00						
Mg	0.68	0.65	0.27	0.35	0.53	-0.26	0.84	0.13	0.79	1.00					
Mn	0.17	0.02	0.11	0.43	0.39	0.23	0.27	0.70	0.40	0.36	1.00				
Na	0.95	0.67	0.06	0.16	0.36	-0.23	0.79	-0.02	0.75	0.68	0.16	1.00			
S	0.66	0.99	0.19	0.15	0.46	-0.19	0.69	-0.21	0.70	0.66	0.02	0.69	1.00		
Si	-0.21	-0.15	0.18	-0.07	0.03	0.22	0.02	0.29	-0.07	0.13	0.30	-0.12	-0.18	1.00	
TP	0.41	0.29	0.26	0.68	0.67	0.26	0.55	0.49	0.73	0.56	0.55	0.49	0.35	0.10	1.00
NPRatio	-0.33	-0.12	-0.05	-0.42	-0.18	-0.09	-0.42	-0.40	-0.51	-0.41	-0.51	-0.40	-0.17	-0.16	-0.82
Alk.	0.72	0.65	0.28	0.22	0.49	-0.30	0.95	-0.03	0.83	0.89	0.31	0.76	0.65	0.13	0.57
Temp.	0.60	0.40	0.22	0.53	0.69	0.11	0.59	0.37	0.75	0.58	0.39	0.62	0.44	-0.06	0.76
Cond.	0.88	0.74	0.18	0.24	0.46	-0.31	0.95	-0.01	0.85	0.89	0.26	0.91	0.75	-0.01	0.56
DO	-0.03	0.28	-0.17	-0.09	0.04	0.07	0.01	-0.17	0.14	-0.11	-0.10	0.00	0.29	-0.12	0.06
Chla	0.43	0.34	0.09	0.47	0.46	-0.01	0.47	0.23	0.65	0.44	0.32	0.48	0.41	-0.13	0.80
Secchi	-0.33	-0.20	-0.21	-0.77	-0.69	-0.37	-0.44	-0.60	-0.61	-0.46	-0.51	-0.39	-0.26	-0.11	-0.92
Photic	-0.34	-0.17	-0.17	-0.77	-0.65	-0.40	-0.42	-0.63	-0.56	-0.41	-0.54	-0.39	-0.23	-0.07	-0.91
pH	0.54	0.44	-0.09	0.09	0.25	-0.32	0.70	-0.02	0.71	0.64	0.23	0.62	0.47	0.05	0.64
TSS	0.33	0.27	0.23	0.52	0.58	0.15	0.50	0.31	0.65	0.47	0.42	0.41	0.32	0.18	0.85
Diatom_SED	-0.34	-0.24	-0.39	-0.28	-0.49	-0.08	-0.27	-0.12	-0.42	-0.35	-0.23	-0.33	-0.27	-0.04	-0.33
Diatom_MO	-0.18	-0.07	-0.28	-0.11	-0.28	-0.20	-0.02	-0.08	-0.09	0.04	-0.05	-0.17	-0.08	0.01	-0.07
Thermo	-0.05	-0.03	0.11	-0.55	-0.39	-0.43	-0.14	-0.45	-0.17	-0.07	-0.28	-0.13	-0.08	-0.11	-0.56
ZStation	-0.26	-0.10	0.22	-0.50	-0.40	-0.28	-0.27	-0.50	-0.35	-0.20	-0.45	-0.34	-0.15	-0.10	-0.70
Volume	-0.06	-0.02	0.44	-0.20	0.06	-0.21	0.05	-0.34	0.02	0.05	-0.25	-0.14	-0.05	-0.03	-0.28
Lake_area	0.01	-0.01	0.50	0.07	0.28	-0.06	0.16	-0.11	0.18	0.15	-0.07	-0.07	-0.02	0.00	0.03
Volume/Lake area	-0.15	-0.01	0.25	-0.53	-0.29	-0.35	-0.12	-0.61	-0.22	-0.15	-0.46	-0.23	-0.05	-0.11	-0.63
Lake area/ZStation	-0.09	-0.05	0.46	-0.16	0.07	-0.17	0.03	-0.27	0.02	0.05	-0.21	-0.18	-0.08	-0.01	-0.24
Volume/ZStation	0.04	0.02	0.46	-0.01	0.24	-0.12	0.17	-0.21	0.17	0.13	-0.13	-0.03	0.01	-0.02	-0.04
Watershed_area	0.10	0.09	0.49	0.10	0.38	0.02	0.26	0.11	0.29	0.29	0.12	0.04	0.07	0.10	0.15
W/L_Area	0.20	0.22	0.13	0.10	0.28	0.16	0.27	0.47	0.30	0.36	0.39	0.20	0.19	0.22	0.26
Qs	0.13	0.15	0.15	0.18	0.32	0.30	0.23	0.57	0.25	0.31	0.42	0.15	0.11	0.36	0.26
Type	-0.27	-0.20	0.11	-0.44	-0.45	-0.27	-0.30	-0.41	-0.39	-0.22	-0.37	-0.34	-0.24	0.04	-0.65
ALT	-0.51	-0.45	-0.25	-0.28	-0.48	0.12	-0.51	-0.04	-0.64	-0.38	-0.09	-0.50	-0.46	0.24	-0.40
LAT	-0.48	-0.40	-0.13	-0.21	-0.40	0.36	-0.58	0.11	-0.52	-0.59	-0.14	-0.38	-0.41	0.23	-0.27
LONG	-0.32	-0.30	0.26	0.31	0.20	0.45	-0.14	0.37	-0.08	-0.08	0.26	-0.25	-0.27	0.38	0.28
STRAHLER	0.13	0.08	0.40	-0.01	0.28	-0.11	0.31	0.13	0.32	0.35	0.23	0.08	0.04	0.17	0.15
GNEISS	-0.34	-0.29	-0.12	-0.27	-0.25	0.28	-0.45	0.16	-0.34	-0.51	-0.02	-0.33	-0.35	0.24	-0.33
FELSIQ	0.23	0.15	-0.26	-0.16	0.06	0.00	0.09	-0.06	0.04	0.05	-0.27	0.30	0.14	-0.17	-0.19
MAFIQ	-0.13	-0.05	-0.14	-0.17	-0.29	-0.22	-0.17	-0.22	-0.30	-0.16	-0.23	-0.19	-0.06	-0.15	-0.33
ARGIL	0.27	0.11	0.27	0.53	0.35	0.03	0.43	0.19	0.43	0.35	0.34	0.19	0.15	-0.23	0.50
CARBO	0.38	0.71	0.19	0.12	0.35	0.00	0.45	-0.14	0.44	0.38	0.03	0.49	0.72	0.09	0.36
SILIC	0.00	-0.05	0.15	0.14	0.29	0.00	-0.11	-0.01	0.01	-0.01	-0.12	0.03	-0.01	-0.21	0.12
INTERM	-0.43	-0.24	-0.14	-0.40	-0.42	-0.11	-0.36	-0.24	-0.34	-0.29	-0.08	-0.35	-0.24	0.18	-0.21
ULTR_MAFIQ	0.12	-0.11	0.10	0.19	0.13	-0.06	0.13	0.18	0.12	0.43	0.24	0.09	-0.08	0.29	0.22
ALLUV	-0.13	-0.05	-0.03	-0.14	-0.11	-0.06	-0.24	-0.02	-0.14	-0.24	-0.11	-0.16	-0.07	-0.25	-0.18
FLUVIO	-0.42	-0.31	-0.14	-0.32	-0.43	0.01	-0.33	-0.13	-0.38	-0.31	-0.04	-0.34	-0.30	0.20	-0.20
LACU	0.01	0.01	-0.20	-0.08	-0.04	-0.06	0.02	-0.08	0.17	0.08	-0.09	0.04	0.01	-0.06	0.06
TILL	0.26	0.20	0.17	0.35	0.33	0.16	0.24	0.17	0.17	0.26	0.11	0.20	0.20	-0.04	0.13
ROCK	0.08	0.01	0.00	-0.05	-0.05	-0.07	0.06	0.09	0.11	0.02	0.26	0.07	0.01	0.20	0.07
MARIN	0.00	-0.05	0.03	0.11	0.20	-0.02	-0.10	-0.08	-0.01	-0.12	-0.19	0.06	-0.01	-0.26	0.12
Ecolog.	-0.05	0.10	-0.15	-0.15	-0.15	0.15	-0.30	-0.19	-0.14	-0.35	-0.39	0.07	0.12	-0.13	-0.10

Note: Significance based on Bonferroni-adjusted probabilities:  $r \geq 0.43$ ,  $p$  value  $\leq 0.05$ ;  $r \geq 0.47$ ,  $p$  value  $\leq 0.01$ .

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by Depository Services Program on 07/31/14  
For personal use only.

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Causative variables					Response variables							Morphometric and geographical variables			
NPRatio	Alk.	Temp.	Cond.	DO	Chla	Secchi	Photic	pH	TSS	Diatom_SED	Diatom_MO	Thermo	ZStation	Volume	Lake_area
1.00															
-0.43	1.00														
-0.56	0.57	1.00													
-0.44	0.93	0.63	1.00												
-0.01	-0.05	-0.02	-0.05	1.00											
-0.69	0.44	0.69	0.51	0.20	1.00										
0.73	-0.42	-0.78	-0.44	-0.11	-0.79	1.00									
0.75	-0.38	-0.76	-0.42	-0.10	-0.76	0.97	1.00								
-0.64	0.70	0.57	0.72	0.23	0.72	-0.55	-0.51	1.00							
-0.71	0.53	0.74	0.48	0.22	0.77	-0.85	-0.81	0.65	1.00						
0.05	-0.39	-0.37	-0.33	0.13	-0.20	0.26	0.21	-0.09	-0.27	1.00					
-0.14	-0.06	-0.09	-0.06	0.12	0.03	0.06	0.03	0.21	0.00	0.79	1.00				
0.43	-0.08	-0.31	-0.12	-0.16	-0.44	0.66	0.68	-0.31	-0.45	0.03	0.06	1.00			
0.62	-0.22	-0.53	-0.30	-0.10	-0.61	0.74	0.79	-0.47	-0.61	0.13	0.04	0.77	1.00		
0.40	0.13	-0.03	-0.04	-0.07	-0.34	0.33	0.39	-0.25	-0.15	-0.27	-0.17	0.65	0.63	1.00	
0.16	0.21	0.24	0.05	-0.01	-0.07	0.00	0.05	-0.07	0.12	-0.37	-0.19	0.42	0.33	0.91	1.00
0.62	-0.05	-0.46	-0.18	-0.01	-0.57	0.70	0.74	-0.36	-0.48	-0.02	-0.03	0.65	0.81	0.76	0.46
0.36	0.10	0.00	-0.07	-0.07	-0.29	0.29	0.35	-0.22	-0.13	-0.25	-0.14	0.65	0.64	0.98	0.93
0.23	0.23	0.17	0.06	-0.01	-0.15	0.09	0.13	-0.11	0.07	-0.37	-0.22	0.44	0.35	0.94	0.96
0.04	0.31	0.37	0.17	0.01	0.00	-0.12	-0.09	0.04	0.17	-0.32	-0.09	0.35	0.23	0.77	0.89
-0.22	0.27	0.35	0.27	0.06	0.12	-0.26	-0.30	0.22	0.14	0.01	0.18	-0.03	-0.14	-0.03	0.05
-0.22	0.24	0.34	0.21	-0.07	-0.01	-0.29	-0.29	0.11	0.16	0.00	0.11	-0.12	-0.13	-0.05	0.01
0.49	-0.25	-0.49	-0.31	-0.18	-0.53	0.67	0.71	-0.41	-0.53	0.09	-0.01	0.65	0.80	0.52	0.25
0.19	-0.46	-0.67	-0.51	-0.04	-0.37	0.42	0.34	-0.38	-0.43	0.46	0.32	0.11	0.13	-0.17	-0.30
0.07	-0.54	-0.47	-0.54	0.16	-0.31	0.25	0.20	-0.34	-0.29	0.32	0.06	-0.12	0.05	-0.32	-0.39
-0.25	-0.02	-0.01	-0.19	0.02	-0.07	-0.24	-0.28	-0.16	0.20	-0.13	-0.10	-0.24	-0.24	0.02	0.14
-0.06	0.37	0.36	0.22	-0.14	-0.03	-0.08	-0.04	0.10	0.15	-0.31	-0.08	0.42	0.23	0.72	0.78
0.21	-0.44	-0.22	-0.47	0.15	-0.36	0.22	0.24	-0.32	-0.20	0.13	-0.13	0.01	0.13	-0.06	-0.14
0.36	0.03	0.07	0.19	-0.03	-0.09	0.16	0.14	0.00	-0.25	0.04	0.05	0.05	0.00	-0.06	-0.11
0.22	-0.25	-0.30	-0.18	-0.19	-0.11	0.25	0.27	-0.17	-0.35	0.18	0.01	0.05	0.29	-0.08	-0.24
-0.41	0.37	0.40	0.34	-0.04	0.38	-0.43	-0.47	0.21	0.40	-0.10	0.13	-0.04	-0.24	0.17	0.36
-0.21	0.47	0.25	0.49	0.25	0.36	-0.23	-0.19	0.32	0.33	-0.31	-0.24	-0.14	-0.20	-0.15	-0.12
0.02	-0.05	0.18	-0.04	0.00	0.07	-0.16	-0.18	-0.03	0.11	-0.10	-0.13	-0.09	-0.13	0.04	0.13
0.01	-0.28	-0.49	-0.33	0.14	-0.16	0.35	0.28	-0.03	-0.18	0.39	0.41	0.10	0.09	-0.16	-0.22
-0.16	0.30	0.12	0.20	-0.23	0.12	-0.22	-0.17	0.20	0.24	-0.38	-0.17	-0.07	-0.07	0.03	0.09
0.17	-0.20	-0.12	-0.22	0.01	-0.11	0.16	0.17	-0.20	-0.20	-0.07	-0.19	0.08	0.13	0.01	-0.06
-0.03	-0.25	-0.51	-0.34	0.14	-0.15	0.32	0.27	-0.11	-0.19	0.37	0.29	0.13	0.12	-0.09	-0.19
-0.12	0.04	0.22	0.04	0.18	0.19	-0.10	-0.09	0.36	0.11	0.03	0.20	0.01	-0.04	-0.06	-0.02
0.09	0.16	0.21	0.24	-0.21	0.03	-0.16	-0.13	-0.06	0.06	-0.18	-0.14	-0.06	0.00	0.00	0.04
-0.14	0.04	0.10	0.05	0.08	0.08	-0.12	-0.14	0.04	0.08	-0.03	0.03	-0.14	-0.17	-0.06	-0.05
-0.05	-0.09	0.17	-0.06	0.06	0.15	-0.19	-0.19	0.04	0.16	-0.03	-0.12	-0.18	-0.20	-0.07	0.01
0.06	-0.32	-0.13	-0.18	0.31	0.07	0.08	0.09	0.00	-0.08	0.12	-0.07	-0.19	-0.05	-0.40	-0.47

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by Depository Services Program on 07/31/14  
For personal use only.

Table A4 (concluded).

	32	33	34	35	36	37	38	39	40	41	42
	Morphometric and geographical variables										
	Volume/ Lake area	Lake area/ ZStation	Volume/ ZStation	Watershed_area	W/L_Area	Qs	Type	ALT	LAT	LONG	Strahler
Cl											
SO <sub>4</sub>											
NO <sub>3</sub>											
COD											
NiT											
Al											
Ca											
Fe											
K											
Mg											
Mn											
Na											
S											
Si											
TP											
NPRatio											
Alk.											
Temp.											
Cond.											
DO											
Chla											
Secchi											
Photic											
pH											
TSS											
Diatom_SED											
Diatom_MO											
Thermo											
ZStation											
Volume											
Lake_area											
Volume/Lake area	1.00										
Lake area/ZStation	0.68	1.00									
Volume/ZStation	0.59	0.91	1.00								
Watershed_area	0.31	0.81	0.84	1.00							
W/L_Area	-0.19	0.00	0.01	0.50	1.00						
Qs	-0.20	-0.02	-0.03	0.41	0.88	1.00					
Type	0.67	0.52	0.28	0.12	-0.21	-0.22	1.00				
ALT	0.09	-0.20	-0.26	-0.31	-0.11	-0.19	0.19	1.00			
LAT	-0.08	-0.30	-0.40	-0.41	-0.15	-0.05	0.10	0.42	1.00		
LONG	-0.17	0.02	0.13	0.14	0.03	0.15	-0.19	0.17	0.47	1.00	
STRAHLER	0.31	0.74	0.75	0.90	0.49	0.42	0.19	-0.29	-0.47	0.03	1.00
GNEISS	0.05	-0.04	-0.15	-0.16	-0.08	0.10	0.25	0.04	0.60	0.13	-0.12
FELSIQ	0.01	-0.09	-0.07	-0.01	0.18	0.11	-0.14	-0.05	-0.07	-0.29	-0.10
MAFIQ	0.15	-0.09	-0.22	-0.25	-0.09	-0.09	0.21	0.13	-0.12	-0.50	-0.14
ARGIL	-0.16	0.18	0.32	0.37	0.13	0.10	-0.27	-0.22	-0.50	0.17	0.31
CARBO	-0.16	-0.17	-0.10	-0.10	0.01	-0.01	-0.22	-0.25	0.03	0.06	-0.11
SILIC	-0.10	0.04	0.11	0.09	-0.05	-0.10	-0.16	-0.08	0.03	0.22	-0.05
INTERM	0.07	-0.15	-0.23	-0.25	-0.13	-0.24	0.05	0.59	0.43	0.24	-0.25
ULTR_MAFIQ	-0.07	0.04	0.06	0.05	-0.04	-0.07	0.02	-0.05	-0.28	0.06	0.10
ALLUV	0.10	0.00	-0.04	-0.08	-0.07	-0.08	0.12	-0.09	0.05	-0.08	-0.01
FLUVIO	0.10	-0.11	-0.16	-0.23	-0.14	-0.23	0.16	0.73	0.55	0.37	-0.21
LACU	-0.11	-0.02	-0.07	0.00	0.04	0.01	0.08	-0.11	0.10	-0.06	0.03
TILL	-0.07	0.03	0.01	0.07	0.08	0.19	-0.11	-0.36	-0.36	-0.27	0.02
ROCK	-0.08	-0.09	-0.01	0.06	0.22	0.13	-0.22	-0.01	-0.09	0.02	0.13
MARIN	-0.16	-0.08	0.01	-0.08	-0.21	-0.21	-0.17	-0.15	0.10	0.16	-0.20
Ecoreg.	-0.13	-0.41	-0.45	-0.52	-0.25	-0.23	0.01	-0.01	0.61	-0.08	-0.55

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by Depository Services Program on 07/31/14  
For personal use only.



43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
Geological variables														
GNEISS	FELSIQ	MAFIQ	ARGIL	CARBO	SILIC	INTERM	ULTR_MAFIQ	ALLUV	FLUVIO	LACU	TILL	ROCK	MARIN	Ecoreg.

1.00														
-0.16	1.00													
-0.23	-0.08	1.00												
-0.46	-0.21	-0.28	1.00											
-0.09	-0.08	-0.10	-0.10	1.00										
-0.17	-0.02	-0.11	-0.07	-0.06	1.00									
-0.06	-0.10	-0.12	-0.19	-0.06	-0.01	1.00								
-0.21	-0.10	-0.12	-0.06	-0.07	0.06	-0.09	1.00							
0.27	-0.05	-0.07	-0.10	-0.03	-0.04	-0.04	-0.05	1.00						
0.03	-0.11	-0.04	-0.21	0.02	-0.03	0.71	-0.10	-0.05	1.00					
0.26	-0.06	-0.07	-0.11	-0.03	-0.04	0.00	-0.05	-0.02	-0.05	1.00				
-0.10	0.18	0.02	0.30	0.05	-0.25	-0.47	0.15	-0.27	-0.64	-0.30	1.00			
-0.12	-0.08	0.28	-0.04	-0.05	0.00	0.06	-0.07	-0.04	0.03	-0.04	-0.35	1.00		
-0.09	-0.05	-0.06	-0.09	-0.03	0.89	-0.04	-0.04	-0.02	-0.04	-0.02	-0.30	-0.03	1.00	
0.41	0.06	0.11	-0.58	0.41	0.13	0.14	-0.35	0.07	0.16	0.29	-0.22	-0.11	0.28	1.00

Can. J. Fish. Aquat. Sci. Downloaded from www.nrcresearchpress.com by Depository Services Program on 07/31/14  
For personal use only.

**Table A5.** List of taxa in the data set with codes associated to their respective authors.

Code	Taxon	Authorities
aamb	<i>Aulacoseira ambigua</i>	(Grunow) Simonsen 1979
acdf	<i>Achnanthidium deflexum</i>	(C.W. Reimer) J.C. Kingston 2000
acop	<i>Amphora copulata</i>	(Kützing) Schoeman & Archibald 1986
adm	<i>Achnanthidium microcephalum</i>	Kützing 1844
admi	<i>Achnanthidium minutissimum</i>	(Kützing) Czarnecki in Czarnecki et Edlund 1995
adnides	<i>Aulacoseira nivaloides</i>	(Camburn) J. English & M. Potapova 2009
aexg	<i>Achnanthes exigua</i>	(Grunow) Cleve and Grunow 1880
afor	<i>Asterionella formosa</i>	Hassall 1850
alir	<i>Aulacoseira lirata</i>	(Ehrenberg) Ross 1986
alri	<i>Psammothidium levanderi</i>	(Hustedt) L.N. Bukhtiyarova 2008
amsa	<i>Achnanthes minutissima</i> var. <i>saprophila</i>	Kobayasi and Mayama 1982
amsc	<i>Achnanthes minutissima</i> var. <i>scotica</i>	(Carter) Lange-Bertalot 1989
aped	<i>Amphora pediculis</i>	(Kützing) Grunow in Schmidt & al. 1875
ariv	<i>Achnanthidium rivulare</i>	Potapova & Ponader 2004
aua1	<i>Aulacoseira alpigena</i>	(Grunow) Krammer 1990
audi	<i>Aulacoseira distans</i> complexe	(Ehrenberg) Simonsen 1979
aufl	<i>Aulacoseira perglabra</i> var. <i>floriniae</i>	(Oestrup) Haworth 1988
auhu	<i>Aulacoseira humilis</i>	(Cleve-Euler) Simonsen
aula	<i>Aulacoseira lacustris</i>	(Grunow) Krammer 1990
auy1	<i>Aulacoseira nygaardii</i>	(Camburn in Camburn & Kingston) Camburn & Charles 2000
aupe	<i>Aulacoseira perglabra</i>	(Oestrup) Haworth 1988
aupu	<i>Aulacoseira pusilla</i>	(Meister) A. Tuji & A. Houki 2004
ausu	<i>Aulacoseira subarctica</i>	(O. Müller) Haworth 1988
aute	<i>Aulacoseira tenella</i>	(Nygaard) Simonsen
auva	<i>Aulacoseira valida</i>	(Grunow) Krammer 1990
avtl	<i>Psammothidium ventralis</i>	(Krasske) Bukhtiyarova and Round 1996
bbre	<i>Brachysira brebissonii</i>	Ross in Hartley 1986
bmic	<i>Brachysira microcephala</i>	(Grunow) Compère 1986
bpro	<i>Brachysira procerca</i>	Lange-Bertalot & Moser 1994
cbac	<i>Caloneis bacillum</i>	(Grunow) Cleve 1894 sensu lato
cbol	<i>Cyclotella bodanica</i> f. <i>lemanica</i>	(O. Müller dans Schröter) Bachmann 1903
ccoc	<i>Cavinula cocconeiformis</i>	(Gregory ex Greville) Mann and Stickle in Round et al. 1990
chme	<i>Chamaepinnularia mediocris</i>	(Krasske) L-B & Krammer in L-B & Metzeltin 1996
cmen	<i>Cyclotella meneghiniana</i>	Cleve et Grunow 1882
cmiga	<i>Cyclotella michiganiana</i>	Skvortzow 1937
cnci	<i>Cymbella neocistula</i>	Krammer 2002
cndi	<i>Cocconeis neodiminuta</i>	Krammer 1991
coce	<i>Cyclotella ocellata</i>	Pantocsek 1901
cplc	<i>Ctenophora pulchella</i>	(Ralfs ex Kützing) Williams & Round 1986
cple	<i>Cocconeis placentula</i> var. <i>euglypta</i>	(Ehrenberg) Grunow 1884
cpse	<i>Cavinula pseudoscutiformis</i>	Mann and Stickle in Round et al. 1990
cpst	<i>Discostella pseudostelligera</i>	(Hustedt) Houk & Klee 2004
csch	<i>Cymbella schubartoides</i>	Camburn et Charles 2000
cste	<i>Cyclotella stelligera</i>	Cleve and Grunow (in Van Heurck) 1882
cyml	<i>Cymbopleura lapponica</i>	(Grunow) Krammer 2003
danc	<i>Diatoma anceps</i>	(Ehrenberg) Kirchner 1878
dite	<i>Diatoma tenuis</i>	Agardh 1812
dmes	<i>Diatoma mesodon</i>	(Ehrenberg) Kützing 1844
dmon	<i>Diatoma moniliformis</i>	Kützing 1833
dova	<i>Diploneis ovalis</i>	(Hilse) Cleve 1891
ebil	<i>Eunotia bilunaris</i>	(Ehrenberg) Mills 1934
eexi	<i>Eunotia exigua</i>	(Brébisson ex Kützing) Rabenhorst 1864
efab	<i>Eunotia faba</i>	Ehrenberg 1838
efle	<i>Eunotia flexuosa</i>	(Brébisson) Kützing 1849
eimp	<i>Eunotia implicata</i>	Nörpel, Lange-Bertalot & Alles 1991
einc	<i>Eunotia incisa</i>	Gregory 1854
emei	<i>Eunotia meisteri</i>	Hustedt 1930
emtr	<i>Eunotia muscicola</i> var. <i>tridentula</i>	Nörpel & Lange-Bertalot 1991
encl	<i>Encyonopsis microcephala</i>	(Grunow) Krammer 1997
enlb	<i>Encyonema lange-bertalotii</i>	Krammer 1997 morphotype 1
enng	<i>Encyonema neogracile</i>	Krammer 1997
enve	<i>Encyonema ventricosum</i>	Grunow in Schmidt 1874
eomi	<i>Eolimna minima</i>	(Grunow) Lange-Bertalot 1998
epec	<i>Eunotia pectinalis</i>	Rabenhorst 1864
epra	<i>Eunotia praerupta</i>	Ehrenberg 1843
esle	<i>Encyonema silesiacum</i>	(Bleisch in Rabenhorst) Mann in Round et al. 1990

Table A5 (continued).

Code	Taxon	Authorities
esub	<i>Eunotia subarcuatoides</i>	Alles Nörpel and Lange-Bertalot 1991
eula	<i>Eucocconeis laevis</i>	(Oestrup) Lange-Bertalot 1999
fcap	<i>Fragilaria capucina</i>	Desmazières 1825
fcro	<i>Fragilaria crotonensis</i>	Kitton 1869
fcrs	<i>Frustulia crassinervia</i>	(Brébisson in W. Smith) Lange-Bertalot and Krammer 1996
fcva	<i>Fragilaria capucina</i> var. <i>vaucheriae</i>	(Kützing) Lange-Bertalot 1980
fexi	<i>Fragilaria exigua</i>	Grunow 1878
fnan	<i>Fragilaria nanana</i>	Lange-Bertalot 1991
fpsm	<i>Frustulia pseudomagaliesmontana</i>	K.E. Camburn & D.F. Charles (2000)
frag1	<i>Fragilaria</i> sp. 1	Lavoie & al. 2008
frho	<i>Frustulia rhomboides</i>	(Ehrenberg) De Toni 1891
fsax	<i>Frustulia saxonica</i>	Rabenhorst 1853
ften	<i>Fragilaria tenera</i>	(W.Smith) Lange-Bertalot
gacc	<i>Geissleria acceptata</i>	(Hustedt) Lange-Bertalot & Metzeltin 1996
gacu	<i>Gomphonema acuminatum</i>	Ehrenberg 1832
gcbc	<i>Gomphonema cymbelliclinum</i>	Reichardt & Lange-Bertalot 1991
gdec	<i>Geissleria decussis</i>	(Hustedt) Lange-Bertalot & Metzeltin 1996
gmic	<i>Gomphonema micropus</i>	Kützing 1844
gpar	<i>Gomphonema parvulum</i>	(Kützing) Van Heurck 1880
gpum	<i>Gomphonema pumilum</i>	(Grunow) Reichardt & Lange-Bertalot 1991
gvar	<i>Gomphonema varistriatum</i>	Camburn & Charles 2000
gyac	<i>Gyrosigma acuminatum</i>	(Kützing) Rabenhorst 1853
hcap	<i>Hippodonta capitata</i>	(Ehrenberg) Lange-Bertalot et al. 1996
kcle	<i>Karayevia clevei</i>	(Grunow) Round & Bukhtiyarova 1996
klat	<i>Karayevia laterostrata</i>	(Hustedt) Kingston 2000
maat	<i>Mayamaea atomus</i>	(Kützing) Lange-Bertalot 1997
mcco	<i>Meridion circulare</i> var. <i>constrictum</i>	(Ralfs) Van Heurck 1880
mcir	<i>Meridion circulare</i>	(Greville) Agardh 1831
audn or mdni	<i>Aulacoseira nivalis</i>	(W. Smith) J. English & M. Potapova 2009
mgra	<i>Aulacoseira granulata</i>	(Ehrenberg) Simonsen 1979
mvar	<i>Melosira varians</i>	Agardh 1827
naan	<i>Navicula angusta</i>	Grunow 1860
naci	<i>Nitzschia acicularis</i>	(Kützing) W.Smith 1853
namp	<i>Nitzschia amphibia</i>	Grunow 1862
navi5	<i>Navicula</i> sp. 5	Lavoie & al. 2008
navs	<i>Navicula sorella</i>	Hohn & Hellermann 1963
ncry	<i>Navicula cryptocephala</i>	Kützing 1844
ncte	<i>Navicula cryptotenella</i>	Lange-Bertalot 1985
ndis	<i>Nitzschia dissipata</i>	(Kützing) Grunow 1862
ndme	<i>Nitzschia dissipata</i> var. <i>media</i>	(Hantzsch) Grunow in Van Heurck 1881
neaf	<i>Neidium affine</i>	(Ehrenberg) Pfitzer 1891 sensu lato
neam	<i>Neidium ampliatum</i> complexe	(Ehrenberg) Krammer 1985
nfon	<i>Nitzschia fonticola</i>	Grunow 1879 sensu lato
nfvi	<i>Neofragilaria virescens</i>	(Ralfs) Williams et Round 1987
ngre	<i>Navicula gregaria</i>	Donkin 1861
niar	<i>Nitzschia archibaldii</i>	Lange-Bertalot 1980
nigr	<i>Nitzschia gracilis</i>	Hantzsch 1860
nimp	<i>Navicula impexa</i>	Hustedt 1961
nipm	<i>Nitzschia perminuta</i>	(Grunow) Peragallo 1903
nlin	<i>Nitzschia linearis</i>	(Agardh) W. Smith 1853
nnot	<i>Navicula notha</i>	Wallace 1960
npad	<i>Nitzschia palea</i> var. <i>debilis</i>	(Kützing) Grunow in Cleve and Grunow 1880
npal	<i>Nitzschia palea</i>	(Kützing) W.Smith 1856
npve	<i>Navicula pseudoventralis</i>	Hustedt 1936
nrad	<i>Navicula radiosa</i>	Kützing 1844
nrch	<i>Navicula reichardtiana</i>	Lange-Bertalot in Lange-Bertalot & Krammer 1989
nrec	<i>Nitzschia recta</i>	Hantzsch in Rabenhorst 1861
nrhy	<i>Navicula rhynchocephala</i>	Kützing 1844
nsem	<i>Sellaphora seminulum</i>	(Grunow) Mann
nshd	<i>Navicula schadei</i>	Krasske 1929
ntrv	<i>Navicula trivialis</i>	Lange-Bertalot 1980
nuif	<i>Nupela impexiformis</i>	(Lange-Bertalot) Lange-Bertalot 1999
nuvi	<i>Navicula vitiosa</i>	Schimanski 1978
nvcc	<i>Navicula viridulacalcis</i>	Lange-Bertalot in Rumrich et al. 2000
nvir	<i>Navicula viridula</i>	(Kützing) Ehrenberg 1838

Table A5 (concluded).

Code	Taxon	Authorities
nzra	<i>Nitzschia radicular</i>	Hustedt 1942
pbio	<i>Psammothidium bioretii</i>	(Germain) Bukhtiyarova et Round 1996
pchl	<i>Psammothidium chlidanos</i>	(Hohn and Hellerman) Lange-Bertalot 1999
pmrg	<i>Psammothidium marginulatum</i>	(Grun) Bukhtiyarova & Round
pprt	<i>Pseudostaurosira parasitica</i>	(W. Smith) Morales 2003
psat	<i>Psammothidium subatomoides</i>	(Hustedt) Bukhtiyarova et Round 1996
psbr	<i>Pseudostaurosira brevistriata</i>	(Grunow in Van Heurck) Williams and Round 1987
psca	<i>Pinnularia subcapitata</i>	Gregory 1856
psct	<i>Psammothidium scoticum</i>	(Flower and Jones) Bukhtiyarova and Round 1996
psdi	<i>Psammothidium didymum</i>	(Hustedt) Bukhtiyarova et Round 1996
ptla	<i>Planothidium lanceolatum</i>	(Brébisson ex Kützing) Round and Bukhtiyarova 1996 sensu lato
ptro	<i>Planothidium rostratum</i>	(Oestrup) Round et Bukhtiyarova 1996
pula	<i>Punctastriata lancettula</i>	(Schumann) Hamilton and Siver (preliminary name)
rsin	<i>Reimeria sinuata</i>	(Gregory) Kociolek & Stoermer 1987
sang	<i>Surirella angusta</i>	Kützing 1844
scon	<i>Staurosira construens</i>	Ehrenberg 1843
scve	<i>Staurosira construens var. venter</i>	(Ehrenberg) Hamilton in Hamilton et al. 1992
sden	<i>Stenopterobia densestriata</i>	(Hustedt) Krammer 1987
seca	<i>Sellaphora capitata</i>	Mann & McDonald in Mann & Al. 2004
sgrl	<i>Stauroneis gracilior</i>	Reichardt 1995
shan	<i>Stephanodiscus hantzschii</i>	Grunow (in Cleve & Grunow)
smed	<i>Stephanodiscus medius</i>	Håkansson 1986
smipar	<i>Stephanodiscus minutulus et parvus</i>	(Kützing) Cleve & Moller 1882 /Stoermer et Håkansson 1984
spin	<i>Stausosirella pinnata</i>	(Ehrenberg) Williams et Round 1987
spup	<i>Sellaphora pupula</i> Complexe	(Kützing) Mereschkowsky 1902
ssam	<i>Stauroneis amphicephala</i>	Kützing 1844
stkr	<i>Stauroneis kriegeri</i>	Patrick
stnh	<i>Stauroneis neohyalina</i>	Lange-Bertalot 1996
tfas	<i>Tabularia fasciculata</i>	(Agardh) Williams & Round 1986
tfen	<i>Tabularia fenestrata</i>	(Lyngbye) Kützing 1844
tflo	<i>Tabularia flocculosa</i>	(Roth) Kützing 1844
tqua	<i>Tabularia quadrisepata (quadrisepata)</i>	Knudson 1952
uuln	<i>Ulnaria ulna</i>	(Kützing) Compère 2001

Note: Codes are listed in alphabetical order.