Journal of Paleolimnology **27:** 151–171, 2002.

WA-PLS diatom-based pH, TP and DOC inference models from 42 lakes in the Abitibi clay belt area (Quebec, Canada)

Mihaela Enache[1,2] & Yves T. Prairie[1]
[1]*Département des sciences biologiques, Université du Québec à Montréal, Case postale 8888, succ. Centre-ville, Montréal, Québec, Canada H3C 3P8 (E-mail: prairie.yves@uqam.ca)*
[2]*Current address: Department of Biological Sciences, Queen's University, Kingston, Ontario, Canada K7L 3N6 (E-mail: enachem@biology.queensu.ca)*

Received 20 December 1999; accepted 20 March 2001

Key words: diatoms, Abitibi, WA-PLS, surficial sediments, total phosphorus, pH, dissolved organic carbon, calibration, ordination

Abstract

The relationship between surficial sediment diatom species and measured environmental variables was explored in lakes from the Abitibi region of western Quebec. Diatom assemblages in 42 lakes were identified and their relationship with measured environmental variables was examined using multivariate statistical methods. Canonical correspondence analysis with forward selection and Monte Carlo permutation tests revealed that the three environmental variables pH, TP and DOC each accounted for statistically significant fractions of the variation in diatom taxa. A training set with 164 modern diatom taxa was used to derive transfer functions for lake-water pH, TP and DOC using weighted-averaging-partial-least-squares (WA-PLS) techniques. The models were developed to infer lake water pH, TP and DOC within the ranges 4–8 for pH, 2.75–30.0 µg l^{-1} for TP, and 2.9–18.5 mg l^{-1} for DOC. These quantitative inference models may now be used to help identify and estimate the effects of natural disturbances on the biogeochemistry of Abitibi lakes during their historical development.

Introduction

Effective aquatic ecosystem management requires the examination of long-term data to assess background conditions, natural variability and the response of aquatic systems to natural or anthropogenic disturbances. Direct contemporary measurements are rarely adequate as they usually cover relatively short-time intervals (rarely more than few decades) that may not allow an accurate description of the magnitude and temporal extent of the effects of these disturbances on aquatic ecosystems. However, lacustrine sediments archive information about the biological history of aquatic systems and indirectly their chemical history as well. Thus, paleolimnological techniques are useful to answer some of the basic questions posed by ecosystem managers (Smol, 1992). Interpreting changes in biotic assemblages preserved in the lake sedimentary profiles based on the present-day ecological characteristics of taxa is a useful way to assess the natural range of variation in a lake over time.

Diatoms are amongst the most widely used groups in quantitative paleoecological reconstructions (Stoermer & Smol, 1999). As part of a larger project examining the effects of historical forest fires on the geochemistry of lakes from the Abitibi region of Quebec, we sought to develop quantitative transfer functions between diatom assemblages and lake chemistry. Although such calibrations already exist for other parts of Canada, the distribution of diatom assemblages and their relationship to limnological characteristics in lakes from the Clay Belt area of western Quebec have never been explored. The first objective of this paper is to explore the major patterns of diatom assemblages and their relationship with 15 measured environmental variables collected from 42 lakes from the Abitibi region

of Quebec. The second aim is to develop calibration models for quantitative inference of some of the main chemical characteristics from fossil diatom assemblages, thereby providing a useful tool for inferring lake chemical changes in relation to past disturbances (Enache & Prairie, 2000).

Study area

The study area is located in western Quebec and covers more than 16,500 km^2 between 48°–50° N and 78° 40′–79° 30′W. Our study sites include 42 lakes with well preserved surficial diatom assemblages (Figure 1). All but three are headwater lakes, so as to exclude the influence of upstream lakes and rivers. The lakes and their catchments did not suffer significant human or natural disturbances during the last 10–15 years. They are generally shallow (1.0–30 m), small (0.01–3.95 km^2), with watershed areas between 0.09–15.08 km^2. Generally they are dimictic (except for the very shallow lakes) with thermal stratification during the summer, and span large gradients in physico-chemical characteristics so that they cover many groups of lakes: acidic to alkaline (pH range from 4.2–8.0), oligotrophic to eutrophic (total phosphorus range from 2.8 µg l^{-1} to 52.0 µg l^{-1}), clear to dark (dissolved organic carbon range from 1.8 mg l^{-1} to 18.5 mg l^{-1}) (Table 1).

The geology, morphology, vegetation and soils of the study area are heterogeneous. These factors contribute to the high variability of biogeochemical characteristics of the study lakes. Geologically, the study area

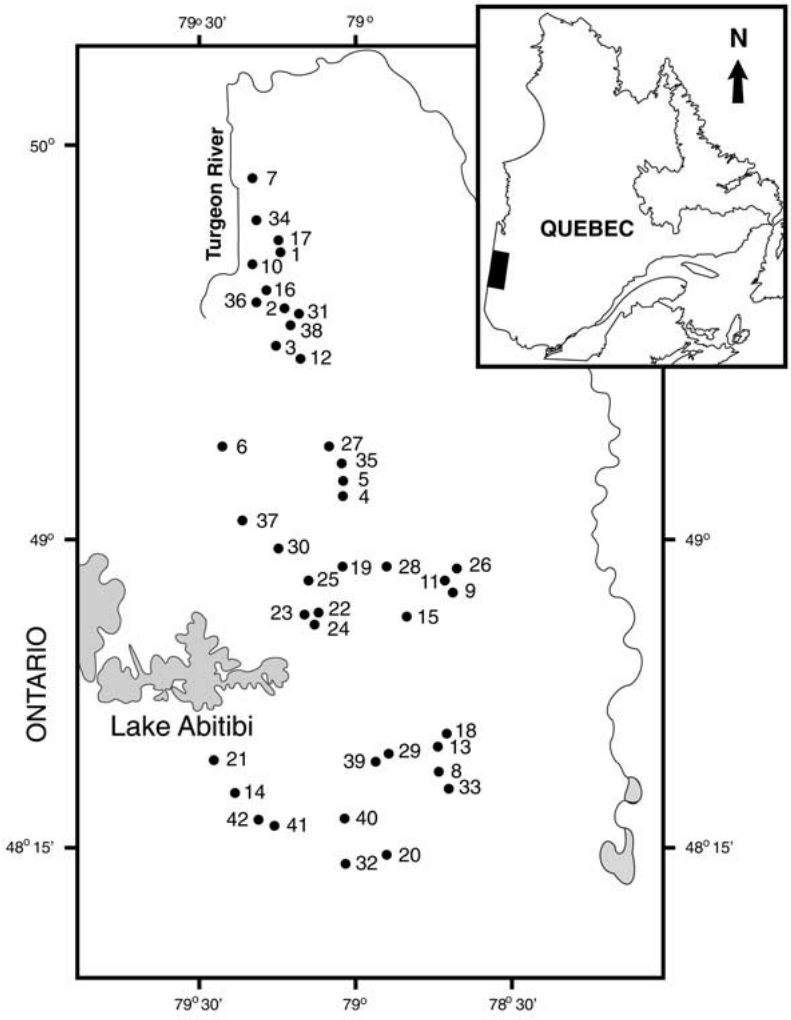

Figure 1. Geographical location of the study lakes. Sample numbers are given in Table 1.

belongs to the Abitibi Subprovince of the Canadian Shield. It is a large folded geosynclinal area from Timmins (Ontario) to Chibougamau, which is made up of volcanic and sedimentary rocks (Commission Géologique du Canada, 1990). Geological substratum of the watersheds surrounding our lakes show considerable petrographical variability: from granitic or granodioritic substrates (lakes KI, Altura, Deux Montagnes, Petit Aldas), intermediate (KJ, Ducharme), to basaltic (Brunet, Pas de Fond, Landry, Francis) or sedimentary substratum (Arsenault, Long, Morillon, Vert). Some lakes (Fabie, Poison) have many types of rocks in their

Table 1. Selected physico-chemical characteristics of 42 sampled lakes, and their minimum, maximum, mean and standard deviation values

No	Lake	pH	TP (μgl^{-1})	TN (μgl^{-1})	DOC (mgl^{-1})	NH_4 (μgl^{-1})	NO_3 (μgl^{-1})	Secchi (m)	Maximumdepth (m)
1	Inconnu 19	6.1	9.5	416.0	15.3	14.0	6.0	1.5	15.0
2	Inconnu 24	8.0	7.0	79.0	2.9	52.5	4.0	4.7	16.0
3	Inconnu 25	6.3	9.0	308.0	8.8	25.5	5.5	3.5	7.8
4	Deux Montagnes	5.9	12.0	424.5	9.7	3.5	3.0	0.8	1.0
5	Inconnu 39	7.4	13.5	213.5	6.4	13.5	2.0	1.6	2.2
6	Altona	5.4	16.0	418.0	11.8	13.0	1.0	1.5	5.0
7	Arsenault	6.1	16.5	501.0	16.7	24.0	4.5	0.8	1.3
8	Brunet	5.5	3.3	202.5	6.6	11.3	10.8	2.0	9.2
9	Debau	6.8	6.8	135.8	8.6	7.5	3.5	2.2	11.8
10	Dieppe	7.1	11.0	466.5	15.8	35.0	10.0	1.4	8.8
11	Disson	7.8	6.3	128.5	4.5	10.0	2.0	2.5	5.0
12	Ducharme	5.4	8.0	342.5	11.0	35.5	6.5	1.6	2.0
13	Edrien	6.3	8.3	217.3	6.7	11.0	2.5	2.0	9.5
14	Fabie	7.3	30.5	327.3	10.9	16.0	2.0	0.3	3.6
15	Pas de Fond	5.8	23.0	511.5	9.9	7.5	3.5	1.7	9.5
16	Garrot	5.2	11.5	376.0	10.1	39.0	1.0	1.9	5.0
17	Gérin	6.8	7.5	328.0	14.6	29.0	35.0	2.0	18.5
18	Highcliff	6.1	3.0	132.5	5.3	8.8	2.5	2.5	7.8
19	Kapekwakata	4.2	9.0	246.0	9.2	31.5	3.0	1.4	10.0
20	KC	4.9	10.5	160.8	8.4	10.3	11.8	1.3	30.0
21	KF (Francis)	5.5	27.5	565.5	14.6	12.0	1.0	0.8	6.1
22	KG	7.4	25.8	303.8	5.7	10.5	1.8	2.8	6.2
23	KH	7.3	17.3	269.3	9.1	12.0	0.5	1.3	7.5
24	KI	6.7	52.0	735.0	18.5	10.0	0.5	1.3	6.1
25	KJ	4.5	3.8	93.0	1.8	6.3	3.3	6.0	10.8
26	KL	4.6	5.0	181.0	8.0	5.5	0.5	1.3	3.8
27	Landry	5.1	7.5	256.5	10.0	5.0	4.5	2.9	8.3
28	Langy	6.4	14.3	224.5	15.0	6.8	2.0	1.0	4.2
29	Lesage	6.1	2.8	152.8	4.7	9.3	5.3	4.0	14.5
30	Leslee	6.6	17.0	593.5	14.0	8.0	5.5	1.5	1.5
31	Long	5.9	5.0	119.0	3.3	30.0	4.5	6.6	28.0
32	Marlon	7.1	22.8	287.0	8.8	16.0	2.8	1.3	2.5
33	Matissard	6.6	17.5	320.3	9.10	12.5	0.8	1.0	2.5
34	Morillon	7.6	26.0	1489.0	14.7	657.0	23.0	1.2	5.5
35	Petit Aldas	5.6	12.5	371.5	13.9	25.5	4.5	1.5	6.0
36	Petoncle	7.0	12.0	501.5	14.1	18.5	1.0	1.3	2.5
37	Poison	5.1	12.0	478.0	16.4	4.0	0.0	1.2	2.0
38	Vert	4.5	5.5	105.0	4.2	25.0	2.0	5.5	14.0
39	Vose	6.0	4.3	158.8	6.05	12.5	8.5	4.5	15.0
40	Waite	6.9	8.8	171.5	6.0	10.8	1.8	1.8	7.2
41	Lassus	6.9	13.3	374.3	13.8	15.5	0.5	1.5	3.5
42	Rich	7.1	6.5	303.8	9.5	11.8	1.0	1.6	5.6
	Minimum	4.2	2.8	79.0	1.8	3.5	0.0	0.3	1.0
	Maximum	8.0	52.0	1489.5	18.5	657.0	35.0	6.6	8.2
	Mean	6.2	12.9	333.1	9.9	31.5	4.6	2.1	30.0
	Standard deviation	1.0	9.4	238.8	4.3	99.5	6.3	1.5	6.5

watersheds (personal observations and geological charts). The surficial geology is very heterogeneous, made of eskers, organic material, and glaciolacustrine deposits from proglacial Lake Ojibway.

Vegetation is primarily represented by coniferous boreal forest. In the south part of the study area (south of 49° N) balsam fire (*Abies balsamea*), white birch (*Betula papyrifera*) and white spruce (*Picea glauca*) are dominant on mesic sites, whereas black spruce (*Picea mariana*), eastern cedar (*Thuja occidentalis*), larch (*Larix laricina*) associated with ash (*Fraxinus nigra*) and elm (*Ulmus americana*) are dominant on bogs and moist sites. Species of poplar (*Populus tremuloides*, *P. balsamifera*) and birch (*Betula papyrifera*) are dominant on sites that were affected by forest fires (Bergeron & Dubuc, 1989). The northern part is dominated by species of spruce (*Picea glauca*, *P. mariana*) and pine (*Pinus banksiana*). Many lakes are surrounded by *Sphagnum* peats (e.g., Ducharme, Francis).

Methods

Each of the 42 lakes were sampled twice during June-August of either 1996 (n = 20) or 1997 (n = 22). Surface sediment samples were collected in the deepest area using a Kajak-Brinkhurst gravity corer (Glew, 1991). The uppermost one cm of sediment was carefully extruded in the field, kept in cool storage and returned to the laboratory for subsequent analyses. These samples were considered to provide an integrated sample (in space and time) of the diatom taxa that accumulated over the previous few years. Measurements of Secchi transparency, temperature and dissolved oxygen, alkalinity and pH were carried out in the field. Integrated surface water samples of the epilimnia were collected in clean polyethylene bottles and immediately transferred to acid-washed and rinsed glass tubes (triplicates) for nutrient analyses. Samples collected for dissolved components were immediately filtered in the field with 0.45 µm Nalgene membrane seringe filters (cellulose acetate filters) and sent to the laboratory in Montreal within 24–48 h. Phosphorus (total and total dissolved) samples were analyzed following potassium persulfate digestion in an autoclave, and total nitrogen and total dissolved nitrogen by alkaline digestion also by persulfate (for more details, see Cattaneo & Prairie, 1995). Chemical analyses were performed on an ALPKEM RFA300 autoanalyser using standard methods.

Preparation, identification and counting of diatom samples

Preparation of diatom samples followed standard methods (Battarbee, 1973, 1986) used for sediment samples without $CaCO_3$. First, a quantity of 0.25 g of dry sediment samples were placed in 10 ml polypropylene tubes and 30% H_2O_2 solution was added to digest organic material. The next day, a strong solution of sulfuric acid (H_2SO_4, 97%) and potassium dichromate ($K_2Cr_2O_7$, M/6) was added to digest the remaining organic material. The digestion process lasted 3–5 days. At the end of the digestion process, the samples were placed in a boiling water bath for 1 hour to complete the digestion. The digested samples were centrifuged and washed with distilled water until the samples were clean and acid-free. The cleaned slurry was resuspended in distilled water (10 ml total volume). A drop (0.2 ml) was added with 1 ml exotic marker (*Eucalyptus globulus* calibrated to 561,133 gr l^{-1}) solution and resuspended in distilled water (5–10 ml final volume). An aliquot (0.2 ml) of this resulting mixture was evaporated onto a coverslip and mounted onto glass slides with Naphrax® (refractive index = 1.65).

For each sample, a minimum of 500 diatom valves was identified and counted with a Leica DMR microscope under oil immersion objectives at 1000X–1600X magnification along random transects. Transects include the peripheral area and only valves including an identifiable central area were counted.

Diatom taxonomy follows standard floras (Hustedt, 1927–1966; Patrick & Reimer, 1975; Germain, 1981; Krammer & Lange-Bertalot, 1986–1991; Lange-Bertalot & Moser, 1994; the PIRLA Diatom Iconograph (Camburn et al., 1984–1986); Lange Bertalot, 1993) and many other references. Because a high taxonomic precision is important in quantitative paleoenvironmental studies (Birks, 1994), diatom taxa were carefully identified to the lowest taxonomic level (i.e., variety and *formae*). Many light micrographs were taken to document the identification. Sediment samples, microscope slides and cleaned diatom slurries are stored at the Department of Biological Sciences, UQAM, Quebec, Canada.

Numerical analyses

Ordination in reduced space allows the derivation of quantitative information on the relationship among des-

criptors (species, environmental variables) and objects (sites) (Legendre & Legendre, 1998). The first step in quantitative environmental reconstructions is to examine whether linear or unimodal methods are appropriate for the available training set in relation to the environmental variables of interest (Birks, 1995). To this end, the biological data set was analyzed by Detrended Correspondence Analysis (DCA) to estimate the length of environmental gradients.

In order to estimate the explanatory power of each environmental variable, a series of Canonical Correspondence Analyses (CCA) was performed using each environmental variable as the sole constraining variable. The statistical significance of each variable was assessed using Monte Carlo unrestricted permutation tests involving 999 permutations (ter Braak, 1990). A CCA (and associated Monte Carlo permutation test) using all measured environmental variables was carried out to find a minimal set of variables that explain, in a statistical sense, some variance within the biological data. Our measured environmental data consisted mostly of physico-chemical and morphological variables (Table 1). CCAs using each group of environmental variables as the sole predictor variables were done to estimate marginal effects of each group of variables. Unique effects of each group of variables and covariances (conditional effects between groups) were estimated using the group of variables as the sole predictor and all other variables as covariables in partial CCAs. Thus, the total variance of diatom taxa was partitioned into fractions representing: (a) the unique and independent contribution of water chemistry, physical and morphological characteristics; (b) conditional effects between the groups of environmental variables; and (c) the unexplained variance (Borcard et al., 1992). CCA is an ordination method appropriate for species abundance and relative frequency data that have a large number of 0 values (absences of taxa) (Legendre & Legendre, 1998). A CCA biplot represents simultaneously the ordination of samples and diatom species and their relationship to environmental variables. In the ordination diagram, environmental variables are represented by arrows, the point of which indicates the direction of maximal quantitative variation.

To eliminate redundancies in the environmental data, groups of significantly ($p \leq 0.05$) correlated environmental variables were identified from a Pearson correlation matrix with Bonferroni adjusted probabilities. Outlier samples (i.e., samples with unusual diatom assemblages, unusual combination of environmental variables or a diatom assemblage with poor relationship to environmental variables) (Hall & Smol, 1992) can strongly affect the predictive power of transfer functions (Birks et al., 1990). Outlier samples were detected by: (1) samples that have extreme (more than 5 times) influence and very high squared residual chi-squared diatance as detected by CCA with the environmental variable as the sole constraining variable (Birks et al., 1990; Lotter et al., 1997); and (2) sample scores that were situated outside the 95% confidence limits about the sample score means in both a PCA of the environmental data and a DCA of the species data (Hall & Smol, 1992). The species data were ln transformed and downweighted for rare taxa.

For the statistically significant environmental variables, diatom-based inference models were developed using weighted-averaging-partial-least-squares (WA-PLS) models (ter Braak & Juggins, 1993). WA-PLS is a unimodal equivalent of the linear-based PLS. This reconstruction procedure has the advantage of taking into account the residual correlations among the biological data, correlations that remain after fitting the environmental variable of interest (ter Braak & Juggins, 1993; Birks, 1995; ter Braak, 1995). In WA-PLS, the first component is selected to maximize the covariance between the vector of weighted averages and the environmental variable of interest. Subsequent components are chosen in the same way but with the restriction that they be orthogonal and hence uncorrelated to earlier components (ter Braak, 1995). The number of components to be retained is determined by cross-validation (leave-one-out-jackknifing) on the basis of prediction error sum-of–squares (PRESS). WA-PLS calibration functions were developed on non-transformed species data.

PCAs, DCAs and CCAs were carried out using CANOCO 3.12 (ter Braak, 1988, 1990); WA-PLS regression and calibration were computed using a SAS/IML (Interactive Matrix Language) implementation of the WA-PLS algorithm (Prairie, unpublished program; available from the authors upon request).

Results and discussion

Environmental variables

A Pearson correlation analysis of ln-transformed environmental variables with Bonferroni adjusted probabilities revealed that many environmental variables are significantly correlated (Table 2). This is not entirely surprising given the broad range of most of the chemical characteristics found in our lakes (Table 1).

Highly correlated pairs ($r > 0.8$) are TP to TDP, TN to TDN, DOC to TN and TDN, alkalinity to pH (Table 2). The relationships among the environmental variables and the position of the lakes with respect to these environmental gradients were further explored using a PCA based on the correlation matrix. Variables with small angles between arrows have high positive correlations (Figure 2). Variables with long arrows have high variance and their proximity to axes determines their relative weight in determining each axis. Lengths of environmental variables' projections were compared to an equilibrium circle of radius $\sqrt{d/p}$ (d: number of axes, p: number of environmental variables, given the hypothesis of an equal contribution to all principal axes; Legendre & Legendre, 1998). Variables reaching further than the equilibrium circle influence are: pH, alkalinity, and TN (Figure 2). The relationship between lake chemistry and watershed characteristics in this region is explored in a separate study (Prairie, unpublished data).

Diatom assemblages

The main characteristic of the diatom assemblages in our lakes is the high species richness. A total of 300 taxa were identified, although only 164 taxa were present in percentages over 1%. This likely reflects the wide range of limnological characteristics covered by the lakes (Table 1). The distributions of the most abundant taxa are illustrated in Figure 3.

Nearly half of the lakes have assemblages with substantial proportions of planktonic species such as *Aulacoseira* spp., *Tabellaria* spp., *Asterionella ralfsii* var. *americana* (e.g., Landry, Debau, Dieppe, etc.). Some lakes are characterized by high densities of *Aulacoseira* species and low species diversity (e.g., KF, Kapekwacata, Marlon, etc.; Figure 3). Other lakes are characterized by high abundances of planktonic *Cyclotella* species, often associated with high percentages of *Aulacoseira* spp. and very diversified benthic forms with low frequencies (lakes Inconnu 19 and Rich).

In other lakes, diatom assemblages are very diverse with dominantly benthic species and lower frequencies of planktonic taxa. *Fragilaria* spp. have relatively high abundances in lakes Inconnu 24, Inconnu 39, Lassus, Waite (generally oligo- to mesotrophic lakes with pH ≥ 7), Arsenault, Leslee (mesotrophic lakes with pH more than 6). In acidic lakes (KC, KL), *F. constricta* var. *stricta* is present in important concentrations associated with *Eunotia* spp. (KC) and especially *Cymbella hebridica*, *Pinnularia braunii* and *Frustulia rhomboides saxonica* (KL). Likely related to local conditions (high ammonia nitrogen concentrations), *Cymbella delicatula* appears only in Lake Morillon in high proportions (21.8%) associated with *Denticula elegans*, *Brachysira vitrea*, and *Achnanthes minutissima* var. *saprophila.*

Table 2. Pearson correlation matrix based on Bonferroni adjusted probabilities of ln transformed environmental variables measured in 42 lake set

Variable	TN	TDN	NO_3	NH_4	TP	TDP	pH	Chla	DOC	Secchi	Alk.	Depth	O_2	L. area	W.
TN	1.000														
TDN	0.960*	1.000													
NO_3	0.017	0.131	1.000												
NH_4	0.290	0.245	0.458*	1.000											
TP	0.728*	0.634*	–0.250	0.151	1.000										
TDP	0.640*	0.633*	–0.156	–0.007	0.875*	1.000									
pH	0.168	0.123	0.046	0.232	0.328*	0.260	1.000								
Chla	0.440*	0.337*	–0.333	–0.041	0.665*	0.496*	0.166	1.000							
DOC	0.848*	0.852*	–0.011	0.108	0.611*	0.615*	0.092	0.333	1.000						
Secchi	–	–	0.213	0.046	–	–	–0.164	–	–	1.000					
Alk.	0.184	0.127	0.050	0.131	0.376*	0.288	0.913*	0.249	0.137	–0.185	1.000				
Depth	–	–	0.429*	0.136	–	–	–0.171	–0.153	–	0.613*	–0.186	1.000			
O_2	0.015	0.012	–0.082	0.034	–0.038	–0.050	–0.150	–	0.106	–0.104	–0.148	–	1.000		
L. area	0.051	0.144	0.334*	0.050	0.045	0.149	0.266	–0.192	0.200	–0.195	0.3128	–0.061	0.419*	1.000	
W.	–0.09	0.00	–0.11	–0.21	0.03	0.26	0.15	–0.169	0.093	–0.254	0.099	–	0.412*	0.512*	1.000

* and **indicate significant correlation at level $p \leq 0.05$ and respectively $p \leq 0.01$; L. = lake, W. = watershed; O_2 represents the hypolimnetic oxygen concentration.

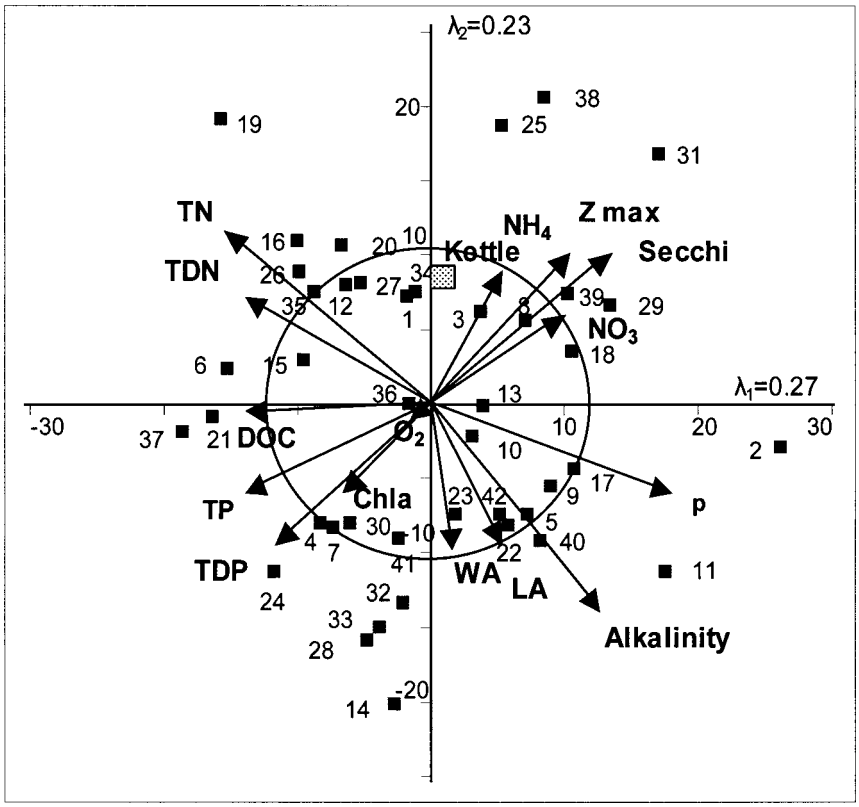

Figure 2. PCA of measured environmental variables of the 42 lake set based on a correlation matrix and the equilibrium contribution circle. Kettle (origin of lake) is a nominal variable. Identification of sample numbers is given in Table 1.

Relationships between diatoms and environmental variables

A DCA of the 164 taxa and seven selected environmental variables resulted in 3.4 standard deviation (SD) units for axis 1 and 2.9 length units for axis 2. Such high values suggest a unimodal response of the taxa in relation to the environmental gradients.

The relationship between diatom taxa and measured environmental variables was explored using CCA. The marginal effect of environmental variables was assessed using each environmental variable as the sole constraining variable. Watershed area, lake area, total dissolved nitrogen (TDN), pH, alkalinity and the nominal variable 'kettle' characterising the origin of the lake, each explain statistically significant parts of the diatom variance ($p \leq 0.05$) as assessed by Monte Carlo unrestricted permutation tests (999 unrestricted permutations). The CCA with forward selection and associated Monte Carlo permutation tests using all environmental variables showed that pH, watershed area, TDN, TDP, maximum depth and NH_4 explain significant portions of diatom variance. All environmental variables explained 44.9% of the total variance of diatom taxa. The lake chemistry variables represented the group of environmental variables that had the most important conditional and marginal effects (Table 3). Because of the high correlations between some chemical variables, TDP, TDN, and alkalinity were eliminated from subsequent analysis in order to minimize redundancies in the environmental variables (Hall & Smol, 1992). Seven environmental variables (pH, TP, TN, DOC, NH_4, NO_3, Secchi) were thus retained for further analysis. These variables are important limnological characteristics. Even though DOC was highly correlated to TN, we retained them both because of their different chemical characteristics.

A CCA between 164 diatom taxa and seven selected environmental variables (pH, TP, TN, DOC, NH_4, NO_3, Secchi) revealed that for the first two axes the eigenvalues were $\lambda_1 = 0.44$ and $\lambda_2 = 0.22$, respectively. The first two axes accounted for 12.3% of the cumulative

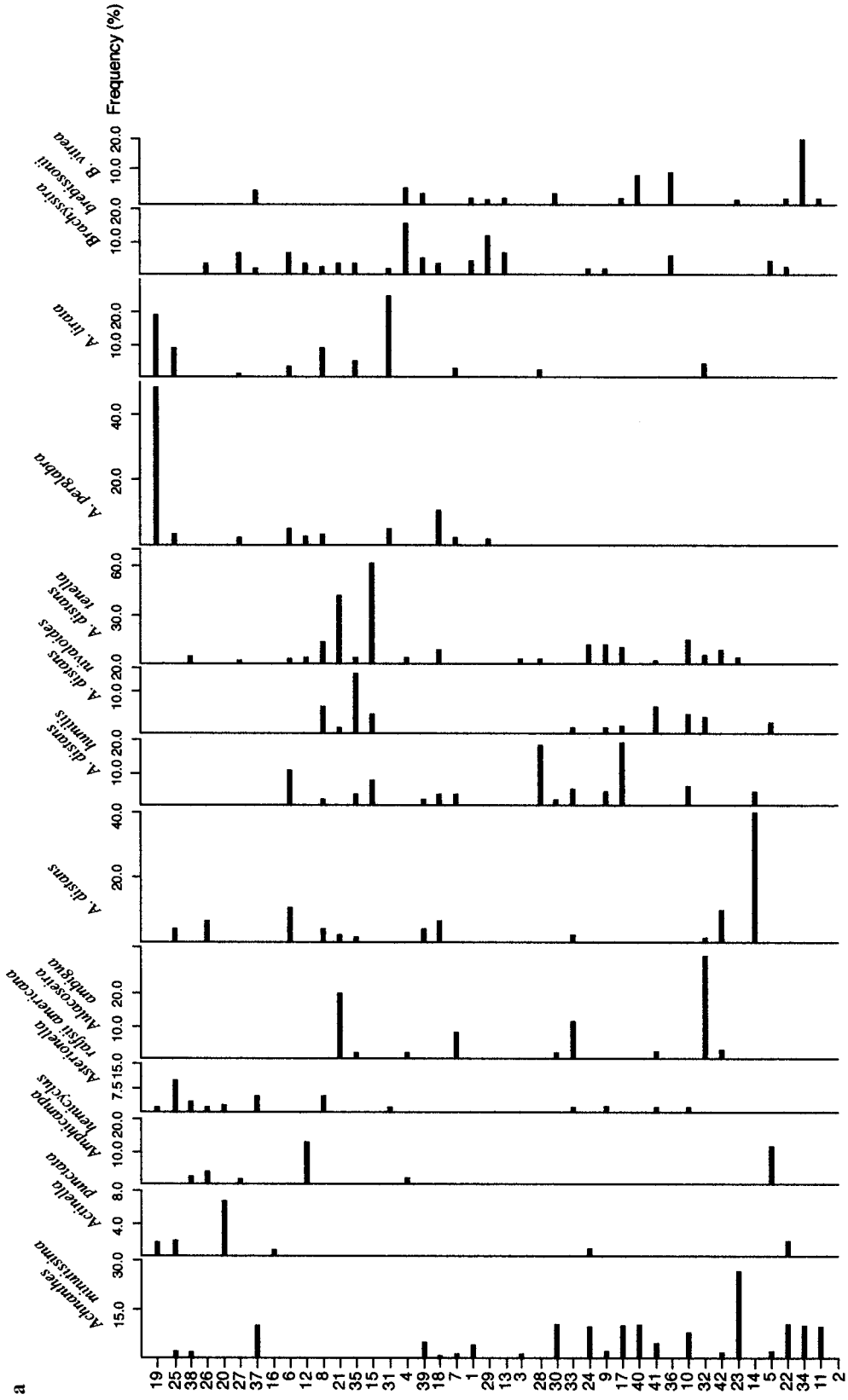

Figure 3. Distribution of selected diatom taxa occurring along the 42 lake set. Sample numbers are given in Table 1. Lakes are ranked from the lowest pH values at the top to the highest values at the bottom of the graph.

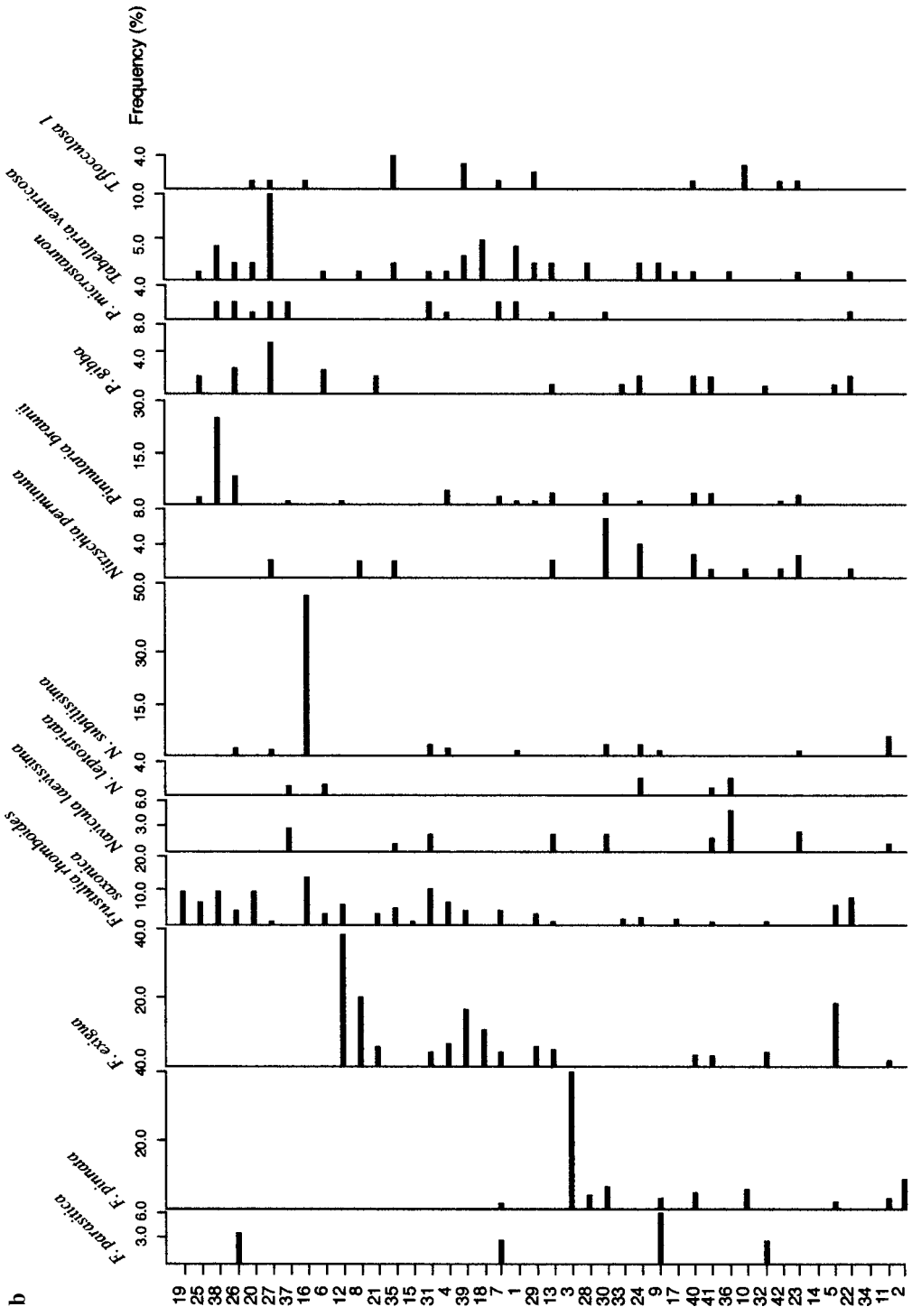

Figure 3. Continued.

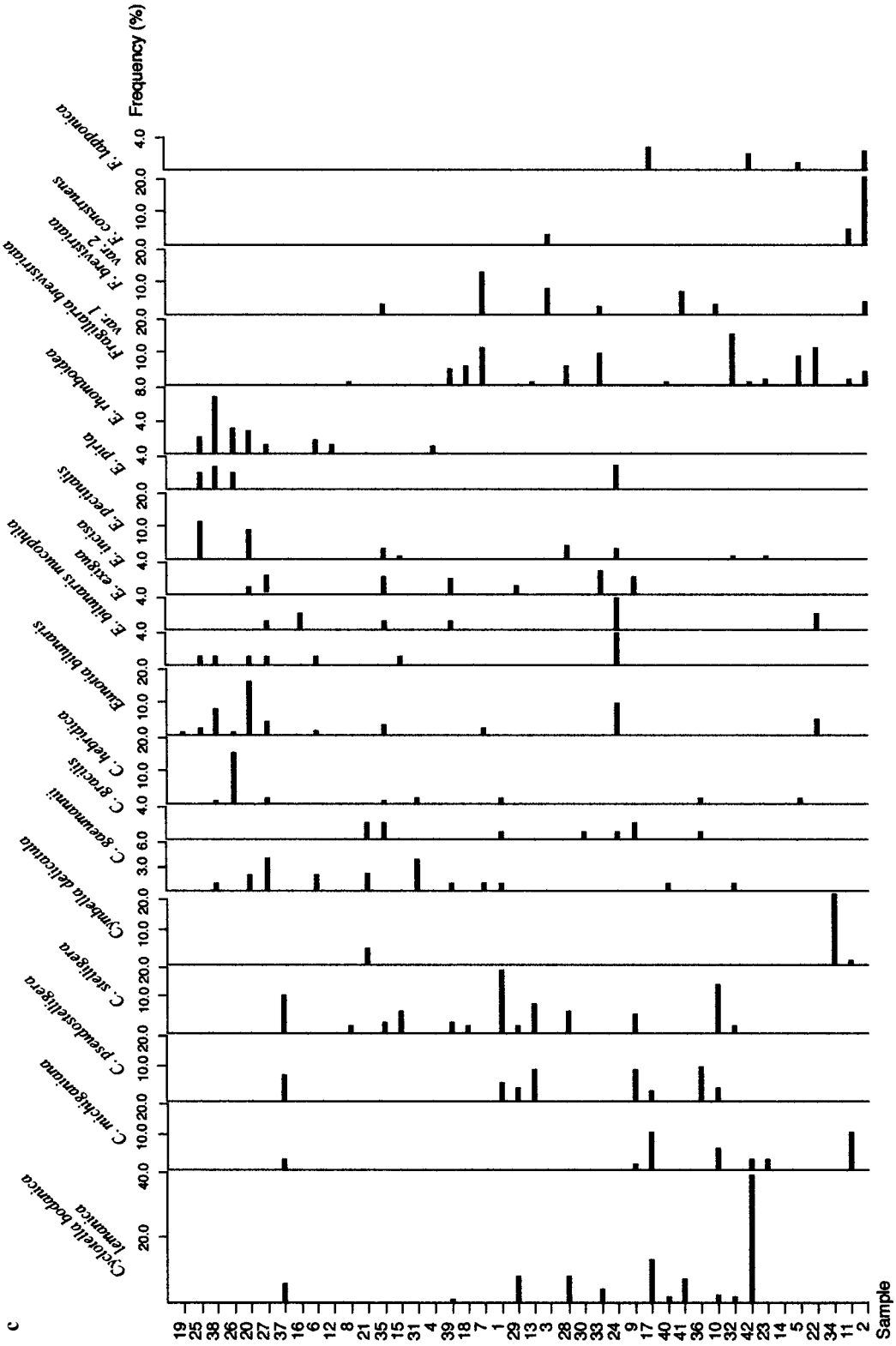

Figure 3. Continued.

Table 3. Unique and marginal effects of groups of environmental variables and partitioning of the total variance in diatom taxa

	Unique	Marginal
Lake chemistry	22.1	26.8
Lake morphology	7.11	15.2
Physical variables	2.53	9.5
Total unique effects	31.74	
Total variance explained (variables and covariables)	44.85	
Total variance explained by covariables	13.11	
Unexplained variance	55.15	

Lake morphology: lake area, watershed area, maximum depth; physical variables: Sechi depth, dissolved O_2 concentration; lake chemistry: TN, TDN, NO_3, NH_4, TP, TDP, pH, DOC, alkalinity; chlorophyll a.

variance in the diatom data. Although this percentage is quite low, such an ordination may still be quite informative (ter Braak, 1990). The first two axes accounted for 48.9% of the variance in the diatom-environment relationships. A high correlation between diatom taxa and the environmental variables for axis 1 (0.93) and axis 2 (0.86) indicates a strong relationship between species and environment. This correlation, illustrated as bi-plots (Figures 4 & 5), shows that environmental variables explain 25.2% of the total variance of species data (= sum of all canonical eigenvalues divided by the sum of all unconstrained eigenvalues × 100 respectively 1.345/5.331 × 100 = 25.2%). The first canonical axis explained 8.2% of the variance in the diatom taxa and the second axis explained 4.1% of the same variance (Table 4).

Forward selection (analogous to the stepwise selection in multiple regression; ter Braak, 1990) identified those environmental variables explaining statistically significant ($p \leq 0.05$) portions of the variation in the diatom data. Monte Carlo permutation tests (999 permutations) were used to test the statistical significance of each forward selected variable. Environmental variables pH ($p < 0.01$), TP ($p < 0.01$), and DOC ($p < 0.05$) each accounted for independent and statistically significant fractions of the variation in the diatom data (Table 5).

The canonical coefficients (Table 6) are the coefficients of the weighted multiple regression of site scores (derived from species scores) to standardized environmental variables. The regression equation of each ca-

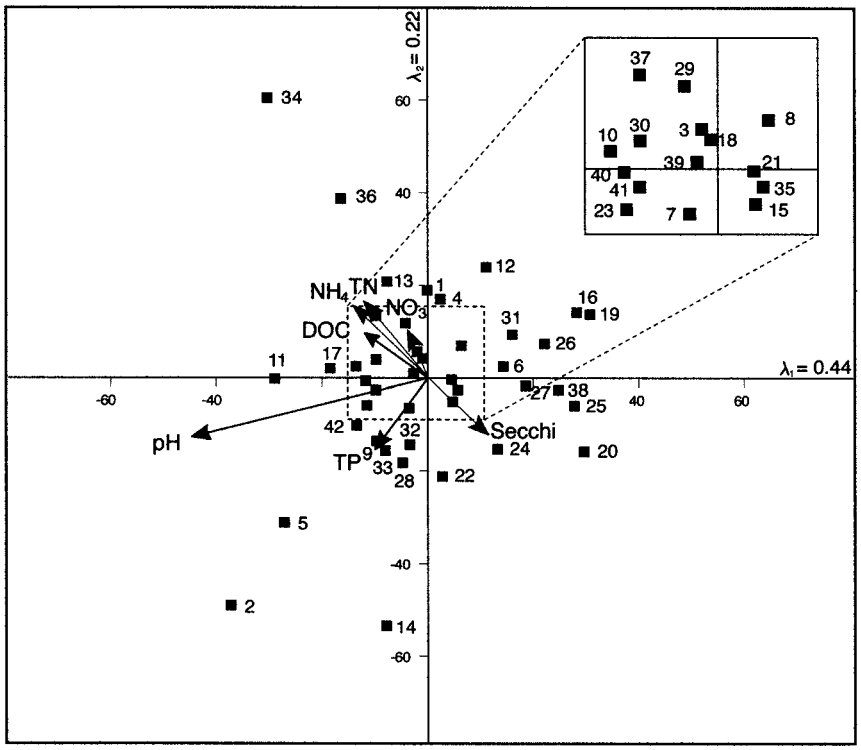

Figure 4. Biplot of environmental variables and sample scores in a CCA of the 42 lake set. Identification of sample numbers is given in Table 1.

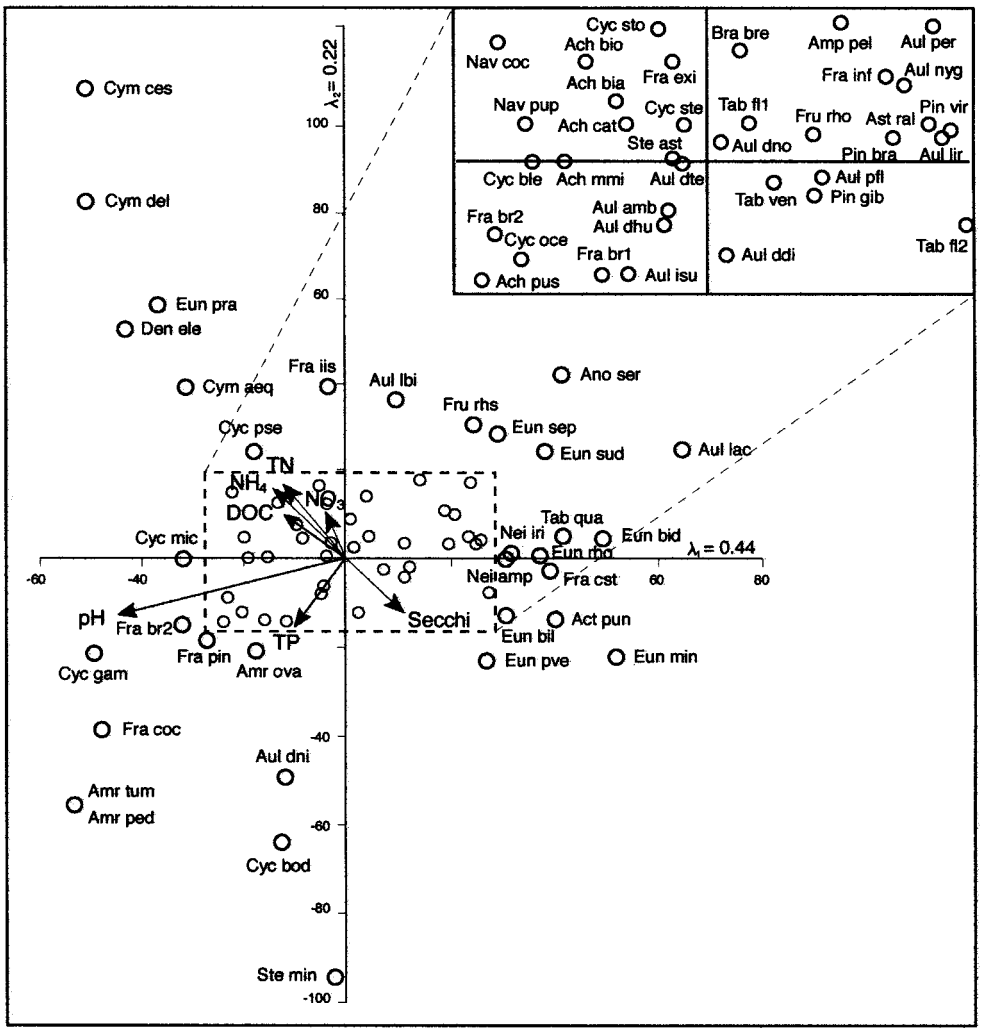

Figure 5. Biplot of environmental variables and species scores in a CCA of the 42 lake set. Full names of species are presented in Table 8. The following species were not cited in Table 7: Ach bia: *Achnanthes biasolettiana* (Kutz.) Grun., Amr ped: *Amphora pediculus* (Kütz.) Grun ex A. Schmidt; Amr thu: *A. thumensis* (A. Mayer) Krieger, Aul lac: *Aulacoseira lacustris* (Grunow) Krammer, Ste min: *Stephanodiscus minutulus* (Kütz.) Round, Fra iis: *Fragilaria inflata* var. *istranii* Hustedt, Eun min: *Eunotia minor* (Kütz.) Grunow.

nonical axis represents the equation of the ecological response of taxa to environmental variables:

$$\text{Axis 1: } -0.7\,\text{pH} + 0.36\text{TP} - 0.08\text{DOC} \quad (1)$$

$$\text{Axis 2: } -0.06\,\text{pH} - 0.60\text{TP} - 0.33\text{DOC} \quad (2)$$

The t-values associated with the canonical coefficients can be used to examine the contribution of environmental variables in describing the species data axes. Thus, pH significantly contributes to axis 1, TP to axes 1 and 2, and DOC to axis 2.

The biplots (Figures 4 & 5) produce weighted-average approximate optimum values of species regarding the environmental variables. These values do not explain all (weighted) variance of species but only a frac-

Table 4. Summary of the CCA with the seven environmental variables (pH, TP, TN, DOC, NH_4, NO_3, Secchi) (first two axes)

Axes	1	2
Eigenvalues	0.44	0.22
Species-environment correlations of species data	0.93	0.87
Cumulative percentage variance	8.2	12.3
Cumulative percentage variance of species-environment relationship	32.5	48.8
Sum of all unconstrained eigenvalues	5.331	
Sum of all canonical eigenvalues	1.345	

Table 5. Statistically significant environmental variables and their associated probabilities

Variable	p	Explained variance	VIF
pH	0.01	0.40	1.24
TP	0.01	0.21	3.19
DOC	0.05	0.16	4.63

999 unrestricted Monte Carlo permutations, adjusted explained variance in diatom taxa and variance inflation factors (VIF), in a CCA with the seven environmental variables (pH, TP, TN, DOC, NH_4, NO_3, Secchi).

tion $(\lambda_1 + \lambda_2) \times 100$/trace (ter Braak, 1987, 1988, 1990). In our case: $(0.44 + 0.22) \times 100/1.345 = 48.9\%$.

Species scores are weighted averages of site (lake) scores. Axis 1 (Figure 4), strongly correlated with pH, contrasts pH-elevated lakes (Inconnu 24, Inconnu 39, Fabie) with acidic lakes (Kapekwacata, Garrot). Axis 2, although weaker, contrasts meso-eutrophic lakes (Kapekwacata, Ducharme) and dark (Petoncle, Dieppe) with clear and oligotrophic lakes (Inconnu 24, Vert, KJ). Both axes are significant (p = 0.00, axis 1 and p = 0.01, axis 2; Monte Carlo permutation test, 999 permutations) and explain 12.4% of the variation in the species data. Species occurring in the upper right quadrant are typical of acidic and/or oligotrophic lakes (*Frustulia rhomboides*, *Aulacoseira perglabra*, *Anomoeoneis serians*, etc.; Figure 5).

In contrast, species that occur in alkaline lakes (e.g., *Fragilaria construens, Amphora pediculus*, *A. tumidula*) are situated on the left part of the biplot. Projections of species to arrows (= environmental variables) approximate the weighted average of species in relation to the respective variable (i.e., their optima). Species projected near the center of the biplot may have their optima in this area, are generalists, or they are not related to the first two axes (Gaillard et al., 1994; ter Braak, 1987).

Calibration: WA-PLS-based pH, TP and DOC inference models

CCA revealed that pH, TP and DOC were the most important environmental variables explaining the distribution of the diatom taxa. We thus decided to develop quantitative inference models for these variables from the diatom assemblages. We used WA-PLS to reconstruct past lake-water pH, TP and DOC from a full data set of 164 diatom species. However, our WA-PLS software allows one to exclude taxa that have broad tolerances (user defined) before they enter the calibration function. Quantitative inferences of pH were derived from the optima of 125 diatom taxa with a maximum tolerance of 2 pH units, while TP and DOC models were derived from, respectively, 88 diatom species (maximum tolerance of diatom species = 5 µg l^{-1} TP) and 70 taxa (maximum tolerance of diatom species = 3 mg l^{-1} DOC). In order to optimize the predictive power of the TP transfer functions, Lake KI was eliminated because of its extremely high TP concentration.

In our case, the WA-PLS algorithm retained the first three components for the pH and TP calibration functions and only the first two components for the DOC calibration (Figure 6).

pH

In our pH inference model, WA-PLS predicted values were close to the measured values in our 42 lakes set (Figure 6). The apparent RMSE of prediction based on a cross-validation (leave-one-out jack-knifing) procedure is 0.25 pH units ($r^2 = 0.93$). Our WA-PLS model did not display any trend in the residuals (Figure 6). The observation that diatoms reflect the pH of lakes is not new, but the predictive power of this calibration model is particularly high. This may reflect the improved statistical fit resulting from the use of WA-PLS but, more

Table 6. Regression coefficients, their t-values and inter-set correlations of environmental variables with axes

Variable	Regression coefficients		t-values		Inter-set correlations	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
pH	–0.694	–0.060	–14.237*	–1.187	–0.867	–0.223
TP	0.362	–0.601	4.619*	–7.337*	–0.160	–0.225
DOC	–0.080	–0.339	–0.847	–3.442*	–0.192	0.153

*t-test significant at $p < 0.05$ in a CCA with the seven environmental variables (pH, TP, TN, DOC, NH_4, NO_3, Secchi).

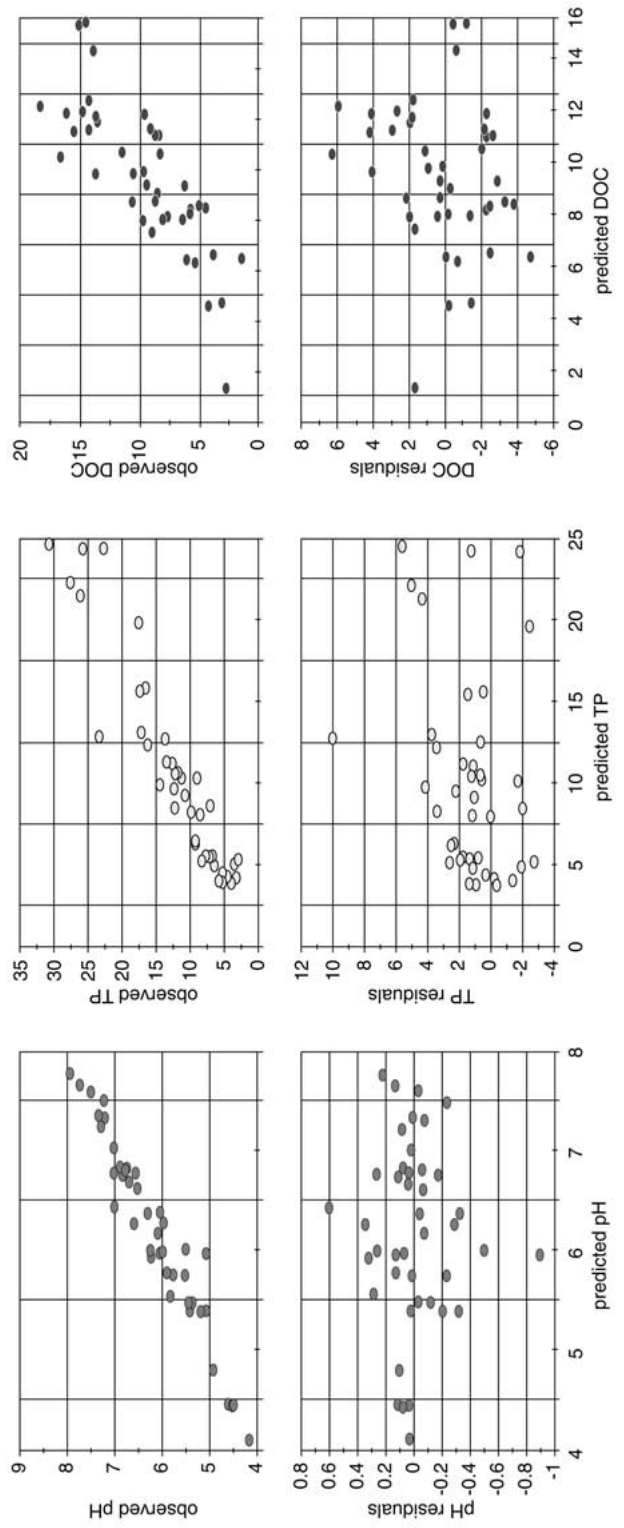

Figure 6. Plots of observed versus diatom-inferred pH, TP and DOC, and predicted versus residual pH (observed-diatom inferred pH), TP (observed-diatom inferred TP) and DOC (observed-diatom inferred DOC) based on WA-PLS calibration models.

likely, is related to the particular chemistry of lake pH in Abitibi, in part related to the glacial sediments. Most of the alkalinity is not derived from the local bedrock, but from carbonates (mostly limestone) transported from the James Bay lowlands and deposited during glacial readvances into proglacial Lake Ojibway in the northern part of the study area (Veillette, 1989). South of N 49°, calcareous rock flour, mostly in the silt fraction, and also derived from the James Bay lowlands, is an important component of the glaciolacustrine clays of Lake Ojibway (Veillette & McClenaghan, 1996). The resulting relationship between pH and alkalinity is that our lakes are relatively more acid for a given alkalinity when compared to other poorly buffered systems on the Canadian Shield. This simple lake pH chemistry may partly explain the surprisingly strong predictive power of the inference model.

Total phosphorus (TP)

A quantitative inference model for total phosphorus was derived from the optima of 88 diatom taxa with maximum tolerance of 5 µg l^{-1} TP. In order to optimize the predictive power of our TP transfer functions, site KI was eliminated because of its extremely high TP concentration relative to the other lakes. The elimination of uninformative taxa with wide tolerances also substantially improved the models. The apparent RMSE of prediction based on the cross-validation procedure was 2.4 µg l^{-1} ($r^2 = 0.89$) for TP, i.e. about 8% of the range in TP values (excluding KI). Again, no trend was observed in the residuals.

Previously published phosphorus inference models from Canadian lakes are few (e.g., Hall & Smol, 1992; Cumming et al., 1994; Reavie & Smol, 2001) and some were developed on ln- or square root-transformed values making their predictive power difficult to compare with ours. Nevertheless, we can calculate arithmetic RMSE values from a Taylor series approximation of the variance of any function (Colquhoun, 1971). This calculation shows that the predictive power of their model and ours is comparable for inferred TP values of about 9 µg l^{-1}. However, for more oligotrophic systems, their model is slightly stronger while the present calibration function is more precise when TP exceeds 10 µg l^{-1} TP. In our data set, about half of the lakes were richer than this threshold. The TP inference model of Reavie & Smol (2001) includes a higher number of lakes with TP values of 20–50 µg l^{-1}, and hence a higher predictive power within this range. Our TP inference model included lakes with TP values up to 30 µg l^{-1}.

DOC

Of the existing Canadian diatom-based DOC inference models (Kingston & Birks, 1990; Pienitz & Smol, 1993; Fallu & Pienitz, 1999), that of Fallu & Pienitz (1999) is both geographically (eastern shore of Hudson and James Bay) and chemically closest to ours. Their mean and range in DOC concentrations are very similar. In spite of this similarity and the same fitting technique used (WA-PLS), the model of Fallu & Pienitz (1999) has greater predictive power with an r^2 of 0.81 and Root Mean Square Error of prediction of 1.23 mg l^{-1}, as compared to the model developed here for Abitibi with an RMSEP of 2.47 mg l^{-1} ($r^2 = 0.65$, Figure 6). In our modelling cross-validation procedure (leave-one-out jackknifing), only two WA-PLS components were retained (compared with 3 for Fallu & Pienitz, 1999). Although relatively less precise than the phosphorus and pH calibration functions, our DOC inference model should still be useful in reconstructing major historical trends in DOC concentrations.

The reasons for the lower precision of our DOC model are not clear. Even the chromophoric nature of the DOC is apparently similar in the two regions. Indeed, DOC is the main factor controlling light penetration and hence water transparency in these lakes (Figure 7). An ANCOVA analysis revealed that the relationship between DOC and Secchi transparency is strong but not statistically different among the regions ($p > 0.05$). This suggests that other factors may modulate the influence of DOC on the diatom assemblages in this region.

Taxa optima

As revealed by CCA, pH appears to be the primary factor influencing the distribution of diatom species in lakes from the Abitibi area. Moreover, most diatom species have narrow pH tolerances. In contrast, diatom species have wider tolerances to TP ranges and particularly to DOC. Species that present well defined optima and narrow tolerances are good indicators and predictors of the environmental variable in question.

Most taxa in the training set had tolerances smaller than 0.5 pH units and most of their optima (Table 7) are similar to data available in the literature (Charles, 1985; Whitmore, 1989; Hickman & Reasoner, 1994; Pienitz et al., 1995). *Eunotia bidentula*, *E. rhomboidea*, *Fragilaria constricta* var. *stricta*, *Tabellaria flocculosa*, *T. quadriseptata*, *Cymbella hebridica* (known as acidophilous taxa) (Charles, 1985; Tolonen et al., 1986; Whitmore, 1989) and *Aulacoseira perglabra* are good

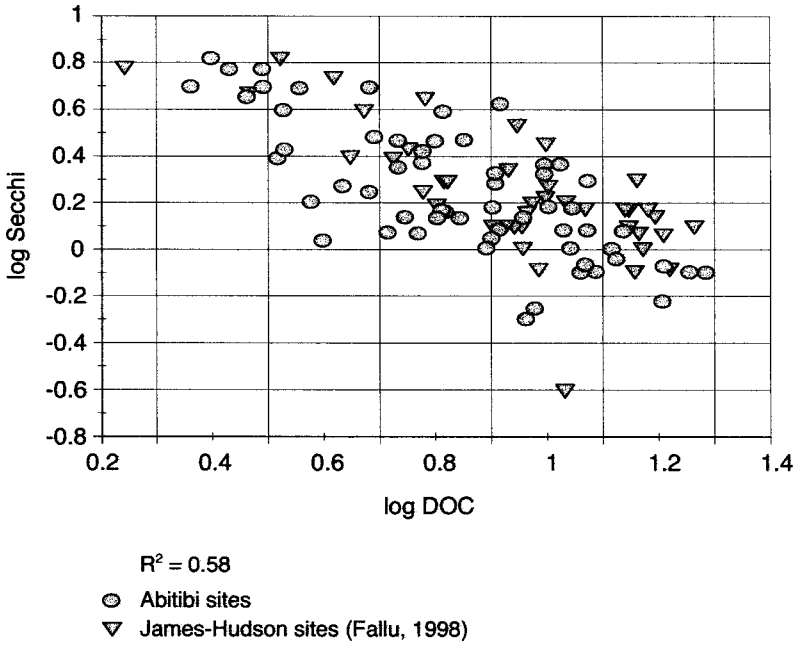

Figure 7. Plot of log Secchi versus log DOC based on a regression analysis in lakes from Abitibi region relative to lakes from James-Hudson region from Quebec (Fallu, 1998).

indicators of pH values less than 5.5 units (Table 7). Other species have their optima near circumneutral conditions: *Cyclotella pseudostelligera*, *Fragilaria brevistriata* varieties 1 and 2. *Cyclotella pseudostelligera* was mentioned as a circumneutral taxon by Stevenson (1989), whereas *F. brevistriata* var. 1 and 2 from Abitibi have slightly lower pH optima as *F. brevistriata* in Charles (1985), who deemed it an alkaliphilous taxon. On the other hand, *Cymbella microcephala*, *C. delicatula*, *Denticula elegans*, *Fragilaria brevistriata* variety 3, *F. construens*, *F. lapponica*, *F. pinnata*, *Navicula diluviana* are alkaliphilous taxa with optima more than 7.5 pH units. Most of these taxa have already been indicated as alkaliphilous by Charles (1985), Tolonen et al. (1986) and Pienitz et al. (1995). Some taxa, such as *Aulacoseira ambigua*, *A. italica* var. *subarctica*, *A. lirata* var. *biseriata*, have pH optima that are slightly lower than indicated by the authors in other areas (Charles, 1985; Gasse, 1986; Tolonen et al., 1986). It is interesting to note that some planktonic species, such as *A. ambigua*, *A. distans* var. *tenella*, *A. distans* var. *nivaloides*, *A. lirata*, *A. perglabra*, and *Cyclotella stelligera*, can exhibit high relative abundances in strongly acidic lakes (Kapekwacata, Brunet, Petit Aldas, etc., Figure 6).

Of the 164 diatom species within the training set, only 88 species had tolerances smaller than 5 µg l^{-1} TP. Among the species related to low TP levels, the most common are *Asterionella ralfsii* var. *americana*, *Eunotia bidentula*, *Cyclotella meneghiniana*, *Aulacoseira perglabra*, *A. perglabra* var. *floriniae* (Table 7), whereas *Navicula seminulum* and *Nitzschia minutula* are found in mesotrophic conditions, and some taxa such as *Aulacoseira ambigua*, *Cymbella delicatula*, *Eunotia implicata*, *E. flexuosa* have optima ≥ 20 µg l^{-1} TP. While our pH optima corresponded well with available data from other areas, there are many differences in the TP optima of our taxa. For example, *Actinella punctata*, *Aulacoseira ambigua* and *Eunotia meisteri* showed higher optima than those of Charles (1985), but were closer to those of Whitmore (1989). *Fragilaria construens*, generally found in meso- to eutrophic waters (Agbeti & Dickman, 1989; Whitmore, 1989; Hall & Smol, 1992), was associated with oligotrophic lakes from Abitibi (TP optima 7.6 µg l^{-1}). Other taxa, such as *C. bodanica* var. *lemanica* and *C. ocellata*, have TP optima much lower than recorded by Wunsam et al. (1995). *Fragilaria brevistriata*, a species recorded in eutrophic and ultra-eutrophic lakes (Whitmore, 1989) or in mesotrophic waters (Hall & Smol, 1992; Reavie & Smol, 2001), is present in our lakes with three morphological varieties with different TP optima: varieties 1 and 2 are present in mesotrophic waters (TP optima 15.05 and 13.7 µg l^{-1}) and variety 3 abounds in

Table 7. pH, TP and DOC WA-PLS optima of diatom species that have narrow environmental tolerances (tolerances less than 2 pH units, less than 5 µg l^{-1} TP and 3 mg l^{-1} DOC

Taxon code	Taxon name	pH optimum	TP optimum	DOC optimum
Ach bio	*Achnanthes bioretii* Germain	6.08	9.32	
Ach cat	*A. catenata* Bily & Marvan	5.99	10.10	
Ach fal	*A. flexella* var. *alpestris Brun*	6.35	12.75	11.04
Ach mmi	*A. minutissima* Kützing *sensu lato*	6.78		
Ach obl	*A. oblongella* Oestrup	5.94	11.99	13.64
Ach pus	*A. pusilla* (Grunow) De Toni	7.03	14.99	8.55
Ach sua	*Achnanthes subatomoides* (Hustedt) Lange-Bertalot	6.51	14.16	
Ach ven	*A. ventralis* (Krasske) Lange-Bertalot	7.45	8.29	6.69
Ach zie	*A. ziegleri* Lange-Bertalot	7.74	9.68	4.35
Act pun	*Actinella punctata* Lewis	5.29		7.81
Amp hem	*Amphicampa hemicyclus* (Ehrenb.) Ralfs	5.58	8.08	9.42
Amr ova	*Amphora ovalis* Kützing	6.8		
Ast ral	*Asterionella ralfsii* var. *americana* Korn	5.16	7.01	
Aul amb	*Aulacoseira ambigua* (Grunow) Haworth	6.47	20.92	11.24
Aul ddi	*A. distans* var. *distans* (Ehrenb.) Kützing	6.55		10.36
Aul dhu	*A. distans* var. *humilis* Cl.- Eul.	6.34		12.64
Aul dni	*A. distans* var. *nivalis* (Smith) Haworth	6.59	9.71	
Aul dno	*A. distans* var. *nivaloides* Camburn	5.97	11.76	
Aul dte	*A. distans* var. *tenella* (Nygaard) Florin	6.09		11.4
Aul isu	*A. italica* subsp. *subarctica* (O. Müll.) Simonsen	6.72		8.5
Aul ist	*A. italica tenuissima* (Grunow) Simonsen	7.00		9.26
Aul lir	*A. lirata* (Ehrenb.) Ross	5.28	8.09	
Aul lbi	*A. lirata* var. *biseriata* (Grunow) Haworth	5.49	5.30	8.18
Aul nyg	*A. nygaardii* Camburn	5.46	6.99	9.30
Aul per	*A. perglabra* (Oestrup) Haworth	4.76	7.98	8.31
Aul pfl	*A. perglabra* var. *floriniae* Camburn	5.73	5.29	4.91
Aul sp1	*Aulacoseira.* sp. 1	5.69		
Aul sp2	*Aulacoseira.* sp. 2	6.00		7.23
Aul ten	*A. tenuior* (Krammer) Grunow	5.70	13.32	
Bra bre	*Brachyssira brebissonii* Ross	5.94	9.40	
Bra pro	*B. procera* Lange-Bertalot	6.30	10.71	8.24
Bra sty	*B. styriaca* (Grunow in Van Heurck) Ross in Hartley	6.45	10.69	13.07
Bra vit	*B. vitrea* (Grunow) Ross in Hartley	6.78		
Cyc bod	*Cyclotella bodanica* var. *bodanica* Grunow	6.89		9.15
Cyc ble	*C. bodanica* var. *lemanica* Grunow	6.72	8.76	10.97
Cyc men	*C. meneghiniana* Kützing	6.63	4.21	5.05
Cyc mic	*C. michiganiana* Sovereign	7.08	8.61	
Cyc oce	*C. ocellata* Pantocsek	6.61	8.16	
Cyc pse	*C. pseudostelligera* Hustedt	6.32	9.76	
Cyc ps1	*C. pseudostelligera* var. 1	7.06		
Cyc ste	*C. stelligera* Cleve & Grunow	6.16	11.10	
Cyc st1	*C. stelligera* var. 1	6.11	9.5	
Cyc sto	*C. stelligeroides* Hustedt	6.25	9.06	14.43
Cym ces	*Cymbella cesatii* Grunow in A. Schmidt	7.57		14.73
Cym del	*C. delicatula* Kützing	7.58	24.97	14.20
Cym gae	*C. gaeumannii* Meister	5.62	9.75	
Cym gra	*C. gracilis* (Rabh.) Cleve	6.24		14.01
Cym heb	*C. hebridica* Grunow ex Cleve	5.11	6.49	8.81
Cym inc	*C. incerta* (Grunow) Cleve	6.38	10.28	
Cym mic	*C. microcephala* Grunow in Van Heurck	6.56		
Cym sil	*C. silesiaca* Bleish ex Rabenhorst	6.32	14.53	14.53
Den ele	*Denticula elegans* Kützing.	7.30		13.28
Dip mod	*Diploneis modica* Hustedt	6.58	13.80	
Dip ova	*D. ovalis* (Hilse)	7.16		9.84
Dip pue	*D. puella* (Schumann) Cleve	6.74	10.85	12.75
Eun bid	*E. bidentula* W. Smith	4.66	7.09	7.16

Table 7. Continued

Taxon code	Taxon name	pH optimum	TP optimum	DOC optimum
Eun bmu	*E. bilunaris* var. *mucophila* Lange-Bertalot & Nörpel	5.63		
Eun bil	*E. bilunaris* (Ehrenb.) Mills var. *bilunaris sensu lato*	5.45	10.65	
Eun exi	*E. exigua* (Brébisson) Rabenhorst	6.13	13.26	
Eun fab	*E. faba* Ehrenb.	5.38	6.25	
Eun imp	*E. implicata* Nörpel & Lange-Bertalot	6.56	21.67	8.22
Eun inc	*E. incisa* Gregory	5.87	9.85	9.21
Eun mei	*E. meisteri* Hustedt	6.79	13.25	15.65
Eun mic	*E. microcephala* Krasske	5.66	11.5	10.42
Eun nym	*E. nymaniana* Grunow (*sensu* Hustedt)	5.58	9.11	
Eun pal	*E. paludosa* var. *paludosa* Grunow	6.48		
Eun pve	*E. pectinalis* var *ventralis* (Ehrenb.) Hustedt	5.31	9.38	
Eun pir	*E. pirla* Carter & Flower	5.08	4.87	
Eun pol	*E. hexaglyphis* Ehrenb.	6.66		
Eun pra	*E. praerupta* Ehrenb.	7.05	14.21	14.20
Eun rho	*E. rhomboidea* Hustedt	4.84	7.48	6.95
Eun rhy	*E. rhynchocephala* var. *rhynchocephala* Hustedt	5.64	9.23	11.40
Eun sep	*E. septentrionalis* Oestrup	5.34	11.88	11.54
Eun sdi	*E. serra* var. *diadema* (Ehrenb.) Patrick	5.33		4.34
Eun ten	*E. tenella* (Grunow)	5.40	8.38	
Fra br1	*Fragillaria brevistriata* Grunow var. 1	6.74		
Fra br2	*Fragillaria brevistriata* Grunow var. 2	6.73	13.7	
Fra br3	*Fragilaria brevistriata* Grunow var. 3	7.96	6.82	3.28
Fra cam	*F. capucina* var. *amphicephala* (Kützing) Lange-Bertalot	6.76	18.62	11.05
Fra con	*F. constricta* Ehrenb.	5.44	12.07	10.52
Fra cst	*F. constricta* f. *stricta* Cleve	4.98	10.97	8.90
Fra cov	*F. construens* f. *venter* (Ehrenb.) Hustedt	6.91	10.93	
Fra coc	*F. construens* f. *construens* (Ehrenb.) Grunow	7.90	7.60	3.56
Fra inf	*F. inflata* Hustedt	5.25	9.09	10.7
Fra lap	*F. lapponica* Grunow	7.24	10.46	7.75
Fra nan	*F. nanana* Lange-Bertalot	6.65		
Fra par	*F. parasitica* var. *parasitica* (W. Smith) Grunow	6.41		
Fra pin	*F. pinnata* Ehrenb.	7.19	12.15	8.11
Fra exi	*F exigua* Grunow	5.89	7.95	8.52
Fru rhs	*Frustulia rhomboides* var. *saxonica* (Rabenhorst) de Toni	5.49		
Fru rho	*F. rhomboides* var *rhomboides* (Ehrenb.) de Toni	4.59	11.46	10.12
Gom gra	*Gomphonema gracile* Ehrenb.	7.28		6.38
Gom par	*G. parvulum* (Kützing) Kützing	6.66		5.86
Gom sub	*G. subtile* Ehrenb.	6.93	12.32	14.02
Gom sp1	*Gomphonema* sp. 1		7.5	14.6
Nav sp1	*Navicula* sp. 1	5.86	11.28	12.0
Nav coc	*N. cocconeiformis* Greg. ex Grev.	6.57	10.44	14.93
Nav. cry	*N. cryptocephala* Kützing	7.31	19.61	8.10
Nav crn	*N. cryptotenella* Lange-Bertalot	7.31	19.26	
Nav dil	*N. diluviana* Krasske	7.95	6.78	3.36
Nav sp2	*Navicula* sp. 2	7.45		
Nav lae	*N. laevissima* Kützing	6.45	12.15	
Nav lep	*N. leptostriata* Joergensen	6.26	13.07	15.48
Nav mes	*N. mediocris* Krasske	6.11	10.17	9.26
Nav sp3	*Navicula* sp. 3	6.71		
Nav pup	*N. pupula* Kützing	6.76	12.08	
Nav rad	*N. radiosa* Kützing	6.87	14.36	9.66
Nav. sed	*N. seminuloides* Hustedt	6.36		8.30
Nav sub	*N. subtilissima* Cleve	5.42	11.08	10.25
Nei amp	*Neidium ampliatum* Krammer & Lange Bertalot	5.20	8.37	
Nei iri	*N. iridis* (Ehrenb.)	5.04	6.66	
Nit fon	*Nitzschia fonticola* Grunow in Van Heurck	6.34		
Nit min	*N. minutula* Grunow	6.48	12.5	

Table 7. Continued

Taxon code	Taxon name	pH optimum	TP optimum	DOC optimum
Pin bra	*Pinnularia braunii* (Grunow) Cleve	5.33	7.98	
Pin gib	*P. gibba* (Ehrenb.)	5.77		9.46
Pin int	*P. interrupta* (W. Smith)	6.0	11.76	11.83
Pin mai	*P. maior* (Kützing)	6.38		
Pin mi1	*P. microstauron* (Ehrenb.) var. 1	5.76		
Pin mi2	*P. microstauron* (Ehrenb.) var. 2	5.63	10.29	
Pin sp1	*Pinnularia* sp 1	4.90	8.21	8.07
Pin vir	*P. viridis* (Nitzsch) Ehrenb.	5.24	7.89	
Sta agr	*Stauroneis anceps* f. *gracilis* Rabenhorst	5.70		
Sta pho	*S. phoenicenteron* (Nitzsch) Ehrenb.	6.01		
Tab ven	*Tabellaria ventricosa* Kützing	5.82	7.99	
Tab fl1	*T. flocculosa* (Roth) Kützing	6.00	9.44	
Tab fl2	*T. flocculosa* (Roth) Kützing strain III *sensu* Koppen	5.22	7.45	8.83
Tab fl3	*T. flocculosa* (Roth) Kützing var. *linearis* Koppen	6.22		
Tab qua	*T. quadriseptata* Knudson	4.97	8.89	8.26

oligotrophic lakes (TP optimum 6.9 µg l^{-1}, see Table 7). *Aulacoseira ambigua*, *Cymbella silesiaca*, *Fragilaria pinnata*, *Navicula cryptocephala*, *N. pupula*, *N. radiosa*, have higher TP optima (sometimes up to 2 fold) compared to the lakes from Ontario in Reavie & Smol (2001), although they cover a wider TP range and the eutrophic lakes are well represented. *Cyclotella stelligera* and *C. michiganiana* have TP optima that agree with observations of Fritz et al. (1993) in oligo-mesotrophic waters.

Fewer species appear strongly related to DOC concentrations. The WA-PLS regression and calibration model indicated that 70 taxa had DOC tolerances less than 3 mg l^{-1}. *Eunotia serra* var. *diadema* and *Achnanthes ziegleri* are most abundant at low DOC concentrations, *Aulacoseira italica* var. *subarctica* and *A. distans* var. *distans* are linked to intermediate DOC concentrations, whereas *Cyclotella stelligera*, *Cymbella cesatii* and *Eunotia meisteri* are abundant in lakes with higher DOC levels (Table 7). Most of Abitibi taxa have higher DOC optima than taxa recorded by Pienitz & Smol (1993) in the Northwest Territories, Canada. However, as noted by Kingston & Birks (1990), DOC optima of the same taxa can differ for different areas. The proximity of the area studied by Fallu & Pienitz (1999) and the overlap in their species list allows for a more systematic test of this hypothesis. When we tried to correlate the DOC optima obtained from our lakes with those reported in Fallu & Pienitz (1999), no clear relationship emerged ($p > 0.05$). This is a further indication that the DOC optima of diatom taxa are not only dependent on the DOC range, but could also be influenced by other unknown environmental mechanisms. Diatom assemblages from Abitibi lakes are very different from those of other nearby regions (Fallu & Pienitz 1999; Philibert & Prairie, unpublished). It is interesting to note that the Abitibi region has a distinct late Holocene history, having been covered for about 2000 years by proglacial Lake Ojibway. It may be that the potential influence of such biogeographical factors has been underestimated.

Conclusions

The high diversity of environmental conditions in the 42 studied lakes from the Abitibi region resulted in a high diversity in diatom assemblages. Canonical Correspondence Analysis revealed that pH, TP and DOC were the principal factors explaining the distribution of the diatom species. Most taxa are strongly related to pH and are good indicators with narrow pH tolerances. Interestingly, lakes with a particularly rich planktonic diatom flora were often acidic (pH ≤ 5.5: lakes Kapekwacata, Francis, Petit Aldas, Brunet, Poison), in contrast to patterns found in other areas where acidity is not natural but anthropogenic (Charles, 1985; Huttunen & Meriläinen, 1986). Nutrient status also played an important role in the distribution of diatom species. A large number of taxa is strongly related to TP concentrations and have tolerances of less than 5 µg l^{-1} TP. Fewer species have narrow tolerances to DOC.

The strong relationship between diatom species and the measured environmental variables allowed us to develop transfer functions for reconstructions of water

pH, TP and DOC in lakes from the Abitibi area or in lakes with similar diatom assemblages. Consequently, it should possible to use these relationships to quantify the effects of catchment disturbances on lake development.

In spite of the relative restricted number of sites (42), the pH inference model has a particularly high predictive power ($r^2 = 0.93$, RMSEP = 0.25). This model (derived from species with tolerances of less than 2 pH units) can be applied to infer lake pH over a broad gradient (4–8 units).

The TP inference model, derived from 88 taxa with tolerances restricted up to 5 µg l^{-1}, also provided a good predictive model ($r^2 = 0.89$, RMSEP = 2.37 µg l^{-1}). Howewer, the unequal distribution of our sites along the TP gradient diminished the predictive power of the model for lakes in the eutrophic range.

The DOC inference model was less powerful ($r^2 = 0.65$, RMSEP = 2.57 mg l^{-1}) in comparison to those developed for pH and TP. Howewer, its predictive power is similar or somewhat lower to other DOC inference models (Dixit et al., in press a, b). Our results showed that the response of diatom taxa from the Abitibi area to DOC is rather different from that for the same taxa of other areas (Fallu & Pienitz, 1999). This may be the result of the particular biogeographical factors proper to be Abitibi area that may influence the distribution of diatom species.

Acknowledgements

The Sustainable Forestry Management Network of Centres of Excellence provided the financial support for this work. We would like to thank J.P. Smol and B.F. Cumming for a profitable stay at PEARL concerning diatom taxonomy and paleolimnological methods. We would like to thank Y. Bergeron for insisting that lakes from Abitibi were worth studying. An older version of this manuscript benefited from discussions and comments from Alex Wolfe. Thanks to Marlène Le Bel, Dany Diotte and Alexandre Soucisse for the assistance in the field and laboratory preparation.

References

Agbeti, M. & M. Dickman, 1989. Use of lake fossil diatom assemblages to determine historical changes in trophic status. Can. J. Fish. Aquat. Sci. 46: 1013–1021.

Battarbee, R. W., 1973. A new method for the estimation of absolute microfossil numbers, with reference especially to diatoms. Limnol. Oceanogr. 18: 647–653.

Battarbee, R. W., 1986. Diatom analysis. In: Berglund, B. E. (ed.), Handbook of Holocene Palaeocology and Palaeohydrology, chapter 26. John Wiley & Sons Ltd. Publishers, 647–653.

Bergeron, Y. & M. Dubuc, 1989. Succession in the southern part of the Canadian boreal forest. Vegetatio 79: 51–63.

Birks, H. J. B., 1994. The importance of pollen and diatom taxonomic precision in quantitative palaeoenvironmental reconstructions. Rev. Paleobot. Palynol. 83: 107–117.

Birks, H. J. B., 1995. Quantitative palaeoenvironmental reconstructions. In Maddy, D. & J. S. Brew (eds), Statistical Modelling of Quaternary Science Data. Quat. Res. Assoc., Cambridge, 161–254.

Birks, H. J. B., S. Juggins & J. M. Line, 1990. Lake surface-water chemistry reconstructions from paleolimnological data. In Mason, B. J. (ed.), The Surface Waters Acidification Programme. Cambridge University Press, Cambridge, 301–313.

Borcard, D., P. Legendre & P. Drapeau, 1992. Partialling out the spatial component of ecological variation. Ecology 73: 1045–1055.

Camburn, K. E., J. C. Kingston, D. F. Charles, D. S. Anderson, J. Ford, P. R. Sweets, F. R. Turner & M. C. Whiting, 1984–1986. In Camburn, K. E., J. C. Kingston & D. F. Charles (eds), PIRLA Diatom Iconograph. PIRLA Unpublished Report Series.

Cattaneo, A. & Y. T. Prairie, 1995. Temporal variability in the chemical characteristics along the Rivière de l'Achigan: how many samples are necessary to describe stream chemistry. Can. J. Fish. Aquat. Sci. 52: 828–835.

Charles, D. F., 1985. Relationships between surface sediment diatom assemblages and lakewater characteristics in Adirondack lakes. Ecology 66: 994–1011.

Colquhoun, D. 1971. Lectures on biostatistics. Clarendon Press, Oxford, 425 pp.

Commission Géologique du Canada, 1990. Regards sur la géologie du Canada. Energie, Mines et Resources Canada, 32 pp.

Cumming, B. F., S. E. Wilson, R. I. Hall & J. P. Smol, 1995. Diatoms from British Columbia (Canada) Lakes and Their Relationship to Salinity, Nutrients, and Other Limnological Variables. Bibliotheca Diatomologica. J. Cramer, Berlin, 207 pp.

Dixit, S. S., W. Keller, A. S. Dixit & J. P. Smol, 2001a. Diatom-inferred dissolved organic carbon provide assessments of past UV-B penetration in Canadian Shield lakes. Can. J. Fish. Aquat. Sci. (In press).

Dixit, S. S., A. S. Dixit & J. P. Smol, 2001b. Diatom and chrysophyte-based inference models for Sudbury (Ontario, Canada) lakes. J. Paleolim. (In press).

Enache, M. & Y. T. Prairie, 2000. Paleolimnological reconstruction of forest-fire induced changes in lake biogeochemistry (Lac Francis, Abitibi, Quebec, Canada). Can. J. Fish. Aquat. Sci. 57: 146–153.

Fallu, M.-A. & R. Pienitz 1999. Modèle de reconstitution de carbone organique dissous à partir des diatomées lacustres, Jamésie-Hudsonie, Québec. Ecoscience 6: 603–620.

Fallu, M.-A., 1998. Distribution des diatomées d'eau douce en Jamésie-Hudsonie, Québec, et développement d'un modèle de reconstitution paléoclimatique. M.A. Thesis, Université Laval, Québec, Qc.

Fritz, S. C., J. C. Kingston & D. R. Engstrom, 1993. Quantitative trophic reconstruction from sedimentary diatom assemblages: a cautionary tale. Freshwat. Biol. 30: 1–23.

Gaillard, M.-J., H. J. B. Birks, U. Emanuelson, S. Karlsson, Lageras & D. Olausson, 1994. Application of modern pollen/land-use

relationship to the interpretation of pollen diagrams-reconstructions of land-use history in South Sweden, 3000-0 BP. Rev. Paleobot. Palynol. 82: 47–73.

Gasse, F., 1986. East African diatoms and water pH. In Smol, J. P., R. W. Battarbee, R. B. Davis & J. Meriläinen (eds), Diatoms and Lake Acidity. Dr W. Junk Publishers, Dordrecht, 149–168.

Germain, H., 1981. Flore des Diatomées: diatomophycées. Société Nouvelle des Éditions Boubée, Paris, France, 444 pp.

Glew, J., 1991. Miniature gravity corer for recovering short sediment cores. J. Paleolim. 5: 285–287.

Hall, R. I. & J. P. Smol, 1992. A weighted-averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia. Freshwat. Biol. 27: 417–434.

Hickman, M. & M. A. Reasoner, 1994. Diatom responses to late Quaternary vegetation and climate change, and to deposition of two tephras in an alpine and sub-alpine lake in Yoho National Park, British Columbia. J. Paleolim. 11: 173–188.

Hustedt, F., 1930–1966. Die Kieselalgen: Deutschlands, Österreichs und der Schweiz unter Berückisichtigung der übrigen Länder Europas sowie der augrendenzenden Meeresgebiete. Teil 1 (1930, pp. 920), 2 (1959, pp. 845), 3 (1961–1966, pp. 816). Ottokoeltz Science Publishers, Koenigstein, Germany. Reprinted 1991.

Huttunen, P. & J. Meriläinen, 1986. Applications of multivariate techniques to infer limnological conditions from diatom assemblages. In Smol, J. P., R. W. Battarbee, R. B. Davis & J. Meriläinen (eds), Diatoms and Lake Acidity. Dr W. Junk Publishers, Dordrecht, 201–212.

Kingston, J. C. & H. J. B. Birks, 1990. Dissolved organic carbon reconstruction from diatom assemblages in PIRLA project lakes, North America. Phil. Trans. R. Soc., Lond. 327: 279–288.

Krammer, K. & H. Lange-Bertalot, 1986–1991. Bacillariophyceae. Süsswasserflora von Mitteleuropa. Band 2(1–4). Gustav Fischer Verlag, Stuttgart. Vols. 1–4. Teil 1 (1986, pp. 876) 2 (1988, pp. 596) 3 (1991, pp. 576) 4 (1991, pp. 437).

Lange-Bertalot, H., 1993. Bibliotheca Diatomologica Band 27. 85 new taxa and much more than 100 taxonomic clarifications supplementary to Süsswasserflora von Mitteleuropa. Vol. 2/1–4. J. Cramer, Berlin, 434 pp.

Lange-Bertalot, H. & G. Moser, 1994. Brachysira. Monographie der Gattung. Bibliotheca Diatomologica, band 29. J. Cramer, Berlin/Stuttgart, 212 pp.

Legendre, P. & L. Legendre, 1998. Numerical Ecology. Elsevier Science, B.V. Amsterdam, 853 pp.

Lotter, A. F., H. J. B. Birks, W. Hofmann & A. Marchetto, 1997. Modern diatom, chironomid, and chrysophyte cyst assemblages as quantitative indicators for the reconstruction of past environmental conditions in the Alps. I. Climate. J. Paleolim. 18: 395–420.

Patrick, R. & C. Reimer, 1975. The Diatoms of the United States. Vol. 2, part 1. Academy of Natural Sciences, Philadelphia, Monograph 13: 1–213.

Pienitz, R. & J. P. Smol, 1993. Diatom assemblages and their relationship to environmental variables in lakes from the boreal forest-tundra ecotone near Yellowknife, Northwest Territories, Canada. Hydrobiologia 269/270: 391–404.

Pienitz, R., J. P. Smol & H. J. B. Birks, 1995. Assessment of freshwater diatoms as quantitative indicators of past climatic change in the Yukon and Northwest Territories, Canada. J. Paleolim. 13: 21–49.

Reavie, E. D. & J. P. Smol, 2001. Diatom-environmental relationships in 64 alkaline southeastern Ontario (Canada) lakes: a diatom-based model for water quality reconstructions. J. Paleolim. 25: 25–42.

Smol, J. P., 1992. Paleolimnology: an important tool for effective ecosystem health. J. Aq. Ecosys. Health 1: 49–58.

Stoermer, E. F. & J. P. Smol, 1999. The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, Cambridge, 484 pp.

ter Braak, C. J. F., 1987. Unimodal models to relate species to environment. Doctoral thesis, University of Wageningen, 152 pp.

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). Tech. Rep. LWA.88.02, GLW, Wageningen, 95 pp.

ter Braak, C. J. F., 1990. CANOCO: a FORTRAN program for CANonical Community Ordination. Microcomputer Power, Ithaca, NY, USA, 35 pp.

ter Braak, C. J. F., 1995. Non-linear methods for multivariate calibration and their use in paleoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. Chemomet. Intell. Lab. Sys. 28: 165–180.

ter Braak, C. J. F. & S. Juggins, 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. Hydrobiologia 269/270: 485–502.

Tolonen, K., M. Liukkonen, R. Harjula & A. Pätilä, 1986. Acidification of small lakes in Finland documented by sedimentary diatom and chrysophycean remains. In Smol, J. P., R. W. Battarbee, R. B. Davis & J. Meriläinen (eds), Diatoms and Lake Acidity. Dr W. Junk Publishers, Dordrecht, 169–200.

Veillette, J. J., 1989. Ice movements, till sheets and glacial transport in Abitibi-Timiskaming, Québec and Ontario. In DiLabio R. N. W. & W. B. Coker (eds), Geological Survey of Canada. Paper 89–90:139–154.

Veillette, J. J. & M. B. McClenaghan, 1996. Sequence of glacial ice flow in Abitibi-Timiskaming: implications for mineral exploration and dispersal of calcareous rocks from the Hudson Bay basin, Quebec and Ontario. Geological Survey of Canada, Open File 3033, Map 1:500 000.

Whitmore, T. J., 1989. Florida diatom assemblages as indicators of trophic state and pH. Limnol. Oceanogr. 34: 882–895.