Journal of Paleolimnology **27:** 97–115, 2002.

Diatoms as indicators of climatic and limnological change in Swedish Lapland: a 100-lake calibration set and its validation for paleoecological reconstructions

Christian Bigler[1] & Roland I. Hall[1,2]
[1]*Climate Impacts Research Centre CIRC, Box 62, SE-981 07 Abisko and Umeå University, Department of Ecology and Environmental Science, SE-901 87 Umeå, Sweden (E-mail: christian.bigler@ans.kiruna.se)*
[2]*Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
(E-mail: rihall@sciborg.uwaterloo.ca)

Received 26 November 2000; accepted 11 March 2001

Key words: Northern Fennoscandia, Abisko, diatoms, lake sediments, pH, temperature, training-set, transfer functions, weighted averaging partial least squares regression (WA-PLS), intra-set cross-validation

Abstract

This study investigated the distribution of subfossil diatom assemblages in surficial sediments of 100 lakes along steep ecological and climatic gradients in northernmost Sweden (Abisko region, 67.07° N to 68.48° N latitude, 17.67° E to 23.52° E longitude) to develop and cross-validate transfer functions for paleoenvironmental reconstruction. Of 19 environmental variables determined for each site, 15 were included in the statistical analysis. Lake-water pH (8.0%), sedimentary loss-on-ignition (LOI, 5.9%) and estimated mean July air temperature (July T, 4.8%) explained the greatest amounts of variation in the distribution of diatom taxa among the 100 lakes. Temperature and pH optima and tolerances were calculated for abundant taxa. Transfer functions, based on WA-PLS (weighted averaging partial least squares), were developed for pH ($r^2 = 0.77$, root-mean-square-error of prediction (RMSEP) = 0.19 pH units, maximum bias = 0.31, as assessed by leave-one-out cross-validation) based on 99 lakes and for July T ($r^2 = 0.75$, RMSEP = 0.96 °C, max. bias = 1.37 °C) based on the full 100 lake set. We subsequently assessed the ability of the diatom transfer functions to estimate lake-water pH and July T using a form of independent cross-validation. To do this, the 100-lake set was divided in two subsets. An 85-lake training-set (based on single limnological measurements) was used to develop transfer functions with similar performance as those based on the full 100 lakes, and a 15-lake test-set (with 2 years of monthly limnological measurements throughout the ice-free seasons) was used to test the transfer functions developed from the 85-lake training-set. Results from the intra-set cross-validation exercise demonstrated that lake-specific prediction errors (RMSEP) for the 15-lake test-set corresponded closely with the median measured values (pH) and the estimations based on spatial interpolations of data from weather stations (July T). The prediction errors associated with diatom inferences were usually within the range of seasonal and interannual variability. Overall, our results confirm that diatoms can provide reliable and robust estimates of lake-water pH and July T, that WA-PLS is a robust calibration method and that long-term environmental data are needed for further improvement of paleolimnological transfer functions.

Introduction

Uncertainty about future climatic change is one of the most serious global environmental concerns, especially in high-latitude and high-altitude regions where the most pronounced greenhouse warming is expected to occur (Weller & Lange, 1999). Unfortunately, the ability to predict future climatic changes continues to be hampered by a shortage of long-term information of past climatic variability. This is particularly the case in subarctic and arctic regions where instrumental meteorological records are generally too sparse and too short (usually < 100 yrs) to predict the magnitude and rate of climatic change on a regional basis (Holmgren & Tjus, 1996). In the absence of direct climatic records, we must rely on paleoclimatic proxy-data to quantify past climatic variability, evaluate results from General Circulation Models (GCMs) and determine biotic responses to changing ecological and climatic conditions (Battarbee, 2000; Felzer et al., 2000; MacDonald et al., 2000).

Amongst the vast amount of proxy-data preserved in lake-sediments, aquatic organisms such as diatom algae are often recognised as promising biological indicators for quantifying past changes in temperature and associated limnological conditions. This is particularly the case in undisturbed regions where community composition, aquatic productivity and habitat availability are regulated closely by climatic factors (e.g., Smol et al., 1991; Lotter et al., 1999; Smol & Cumming, 2000). However, the relationship between diatom distribution and temperature remains poorly understood and therefore diatom-based temperature transfer functions continue to be met with scepticism (Anderson, 2000). In addition, there are many potentially strong physical (habitat), chemical (pH, nutrients) and biological (food-web effects) influences on diatom assemblages that may act independently of water- or air-temperature (Battarbee, 2000). In contrast, chironomid transfer functions are widely believed to provide more reliable transfer functions for reconstructing air-temperature (Battarbee, 2000), in part because air temperatures directly regulate the adult, reproductive stage. Unfortunately, there has been a paucity of studies designed to assess rigorously and independently the performance of diatom-temperature transfer functions.

Despite uncertainty concerning the performance of diatom-temperature transfer functions, several such transfer functions have been developed from training-sets in high latitude - (Pienitz et al., 1995; Weckström et al., 1997; Rosén et al., 2000a; Bigler et al., 2000) and high altitude-regions (Vyverman & Sabbe, 1995; Wunsam et al., 1995; Lotter et al., 1997). In fact, this development has been encouraged because performance statistics of the diatom-temperature transfer functions are usually comparable to those based on other proxy-indicators (e.g., chironomids, pollen) from the same training-sets (e.g., Lotter et al., 1997; Rosén et al., 2000a). Moreover, a recent multi-proxy paleoecological study demonstrated that quantitative estimates of past temperature changes from diatoms agreed well with those based on chironomids, pollen and near-infrared spectroscopy (NIRS) during most of the Holocene (Rosén et al., 2000b). With such promising results, the onus is now clearly on paleolimnologists to develop methods to rigorously assess the performance of diatom-temperature transfer functions.

Most of the diatom-temperature training-sets are from relatively remote areas to take advantage of low human impacts and to reduce confounding influences of eutrophication, atmospheric pollutants and watershed disturbances that may alter diatom-climate relationships. These remote regions are undoubtedly the best places to develop paleo-temperature transfer functions (Lotter et al., 1999). However, logistical constraints often limit the ability of the investigator to obtain more than a single measurement of environmental variables (e.g., water chemistry, water temperature) and to assess their reliability in a rigorous and independent manner. It is often prohibitively expensive to sample extra lakes that can be used as a test set for independent cross-validation. Moreover, because long-term meteorological records are generally sparse and short, direct validation of diatom-temperature transfer functions via comparison against direct long-term meteorological measurements is often impossible in remote regions.

The present study explores relationships between diatoms and present-day conditions in 100 lakes situated across steep climatic gradients and ecotonal boundaries in Swedish Lapland and develops transfer functions for quantifying past changes in mean July air temperature and pH. We then assess the predictive ability of the transfer functions using a method of independent cross-validation (in addition to bootstrapping or jack-knifing numerical re-sampling approaches that are commonly used to derive performance statistics of weighted-averaging regression and calibration models; *sensu* ter Braak et al. (1993), Birks (1998)). The method uses 'intra-set cross-validation', where the full 100-lake training-set is divided into two parts: an 85-lake training-set that is used to generate the transfer functions and a 15-lake test-set that is used to assess the ability of these transfer functions to estimate the actual, or measured, values. The 85-lake training-set is based on limnological data collected during a single visit to each lake, whereas limnological variables in the 15-lake test-set are based on repeated monthly measurements during the ice-free seasons of two successive years. Especially for the pH transfer function, the intra-set cross-validation allows us to assess the predictive ability of transfer functions that are based on single-point measurements within the context of seasonal variability exhibited by each of the test-set lakes.

Description of study sites

The study area (67.07° N to 68.48° N latitude, 17.67° E to 23.52° E longitude) includes 100 lakes located in the boreal and alpine zones of Swedish Lapland, and extends between 100–300 km north of the Arctic Circle (Figure 1). The area is characterised by steep gradients of temperature and precipitation due to influences of elevation and the Atlantic Ocean, respectively. Annual precipitation varies from > 1000 to 500 mm at the west and east ends, respectively, with a minimum of 304 mm in the central Abisko valley due to orographic effects (Alexandersson et al., 1991). Elevation of sites ranges from 169–1183 m a.s.l., corresponding to estimated mean July air temperatures (July T) of 7.0–14.7 °C (Table 1). The ice-free season ranges from ~ 6 months at low elevation lakes to ≤ 2 months at the highest elevations. A relatively strong linear correlation ($r^2 = 0.58$) exists between surface-water temperature and July T (Larocque et al., in press), similar to patterns observed in other regions (e.g., Livingstone & Lotter, 1998).

Training-set lakes are generally small (< 20 ha), shallow (average maximum depth = 6 m) and oligotrophic headwater lakes. The pH ranges from slightly acidic to slightly alkaline (5.8–8.1). Additional limnological and geographic characteristics are summarised in Table 1.

Catchment vegetation ranges from conifer-dominated (*Pinus sylvestris*, *Picea abies*) boreal forests at low elevations, through mountain-birch woodland (*Betula pubescens* ssp. *tortuosa*) and alpine meadows, to areas with sparse alpine-tundra vegetation (mainly lichens and mosses) or barren ground. The tree-limit

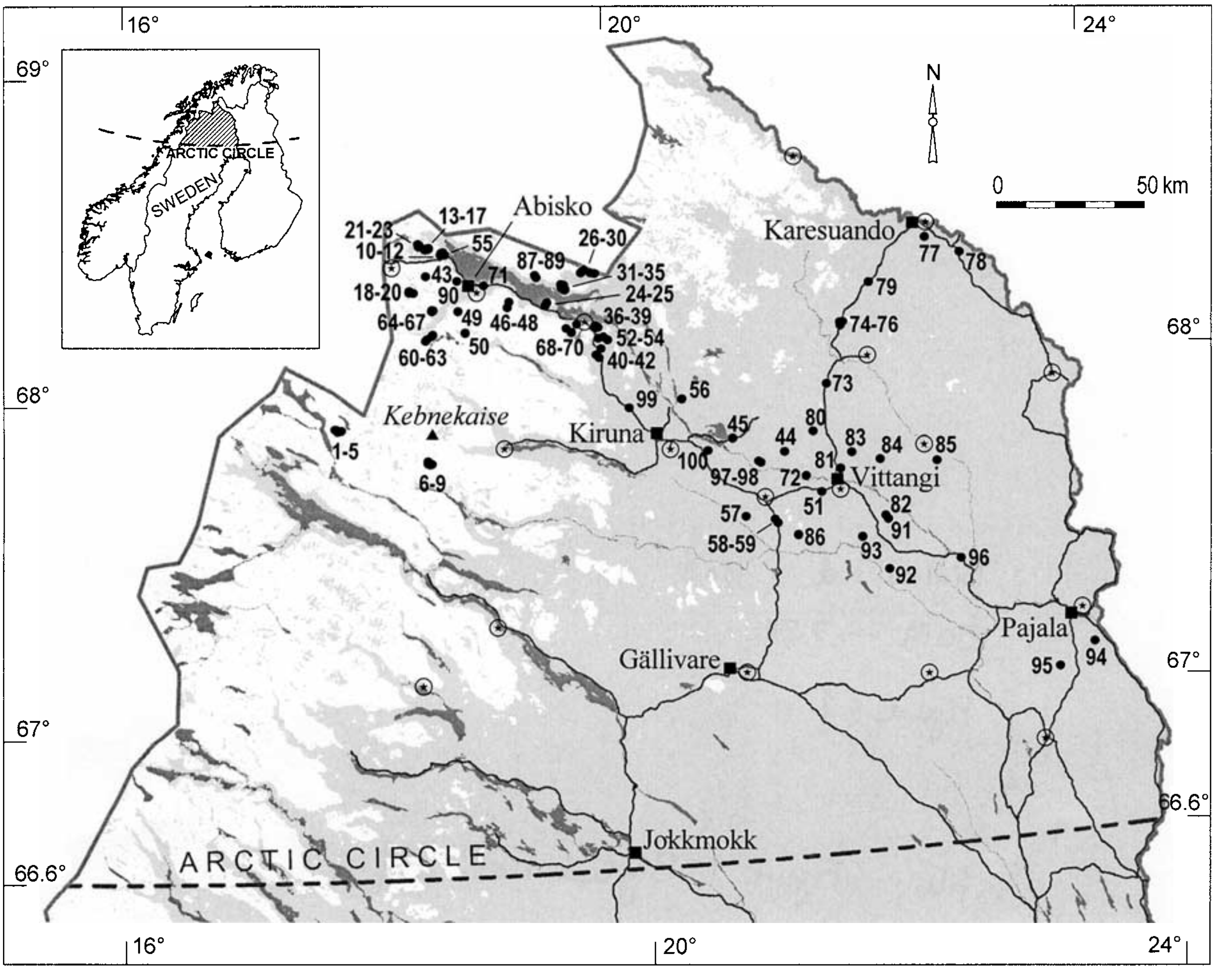

Figure 1. Map of northern Sweden showing locations of the 100 sampled lakes. The locations of the 18 weather stations used for spatial interpolation of the mean July air temperature at each lake are indicated as stars. The forested areas (grey) are distinguised from sites above the tree-limit (white).

Table 1. Summary of the environmental characteristics of the 100 lake-set used as possible explanatory variables for the distributions of surface-sediment diatoms

Nr	Name	Variable	Max	Min	Mean	Median	S.D.
1	Water T	Measured water temp. (°C)	17.5	2.4	10.4	10.5	2.5
2	July T	Estimated mean July air temp. (°C)	14.7	7.0	10.8	11.2	1.9
3	Jan T	Estimated mean January air temp. (°C)	–12.9	–16.4	–14.6	–14.5	0.8
4	Depth	Maximum lake depth (m)	17	2	6	5	3
5	LOI	Loss On Ignition (%)	87.31	2.88	42.17	42.42	16.99
6	pH	pH	8.07	5.79	–	6.72	0.43
7	DOC	DOC (mg/L)	13.4	0.2	3.5	2.8	2.4
8	Si	Si (mg/L)	3.060	0.010	0.789	0.715	0.684
9	Cond	Conductivity (μS/m)	12.800	0.560	2.080	1.640	1.654
10	Ca	Ca (meq/L)	1.030	0.010	0.099	0.066	0.122
11	Mg	Mg (meq/L)	0.175	0.007	0.040	0.027	0.037
12	Na	Na (meq/L)	0.107	0.011	0.036	0.032	0.015
13	K	K (meq/L)	0.050	0.002	0.008	0.006	0.007
14	Sulphate	SO_4 (meq/L)	0.361	0.009	0.041	0.028	0.052
15	Cl	Cl (meq/L)	0.104	0.008	0.025	0.021	0.015
	lat	Latitude	68.48	67.07	68.10	68.20	0.32
	long	Longitude	23.52	17.67	19.82	19.57	1.46
	elev	elevation (m a.s.l.)	1183	169	595	500	260
	secchi	secchi depth (m)	11.1	1.0	4.8	4.5	2.2

Numbered variables were included in the canonical correspondence analyses (CCAs).

(c. 650 m a.s.l.) is the main ecological feature and is formed by mountain birch.

Methods

Field and laboratory analyses

Surface sediments were collected from the deepest part of each training-set lake using a modified Kajak-gravity corer (Renberg, 1991). The deepest part of each lake was located with a portable echo sounder. Diatom slides were prepared following standard methods (Battarbee, 1986; Renberg, 1990). At least 400 diatom valves were counted from each sediment sample at 1000×-magnification using phase contrast. Diatom taxonomy followed the guidelines of the Surface Water Acidification Programme (SWAP) and Krammer & Lange-Bertalot (1986–1991); additional literature was used for the genus *Aulacoseira* (Camburn & Kingston, 1986) and for region-specific diatoms (Hustedt, 1942).

Sediment organic matter content was estimated using standard methods as the percent weight loss-on-ignition (LOI) after heating dried sediment at 500 °C for 1 h (Dean, 1974; Heiri et al., 2001).

Water samples were collected at 1-m water depth from each training-set lake immediately before the sediment samples were obtained. Water-chemistry analyses were performed at the Swedish University of Agricultural Sciences (SLU) in Uppsala. The 15 test-set lakes for intra-set cross-validation were sampled by our limnological colleagues at the Climate Impacts Research Centre (CIRC) at monthly intervals during the ice-free periods of 2 years using identical methods, and water chemistry analyses were performed by the Marine Research Centre, Umeå University. The test-set lakes were distributed evenly over the entire altitudinal gradient in order to cover the full range of ecological and climatic conditions and lake-types.

Mean July and January air temperatures were estimated by spatially interpolating records of Climate Normals (1961–1990) from 18 nearby situated weather stations and applying a lapse rate of 0.57 °C per 100 m elevation (Laaksonen, 1976). These calculations were identical to similar studies that developed diatom- and chironomid-based temperature transfer functions in Fennoscandia (Olander et al., 1999; Brooks & Birks, 2000; Rosén et al., 2000a).

Data analyses

Diatom data were expressed as relative abundances (i.e., % of total diatoms). Species were retained in numerical analyses if they were encountered in at least three training-set lakes with an abundance of > 1% in at least one lake. For all analyses, diatom data were square-root transformed in order to stabilise variances and optimise the ‘signal’ to ‘noise’ ratio in the data (e.g., Brooks & Birks, 2000).

Chemical variables (with exception of pH) were transformed using ln(x+1), whereas other environmental variables were normally distributed and were not transformed. Secchi depth and total organic carbon (TOC) were measured at each training-set lake, but were removed from the environmental data set prior to all statistical analyses. Secchi depth was removed because the Secchi disk was visible on the bottom of most lakes, due the shallowness of most of the sites. We selected dissolved organic carbon (DOC) in place of TOC because they were highly correlated and because DOC contributed almost all of the TOC in the training-set lakes (Davis et al., 1985).

Detrended correspondence analysis (DCA; Hill & Gauch, 1980), with detrending by linear segments and non-linear re-scaling of axes, was used to explore the main patterns of taxonomic variation among the training-set lakes and to estimate the compositional gradient lengths along the first few DCA axes.

The forward-selection option of canonical correspondence analysis (CCA) was used to identify a minimal environmental data set that could explain the variation in diatom distributions among the training-set lakes almost as well as when all environmental variables were used. Statistical significance of each forward selected variable was tested by a Monte Carlo permutation test (999 unrestricted permutations), using a Bonferroni-type adjustment for significance levels as described by Lotter et al. (1997). The strength of the relationships between environmental variables and diatom data was assessed using a series of partial and constrained CCAs with one environmental variable entered at a time. The ratios λ_1/λ_2 provided a measure of the predictive ability of each variable. In general, quantitative inference models can be developed for variables with a high λ_1/λ_2 ratio (Olander et al., 1999). Detrended CCA (DCCA), with detrending by linear segments and non-linear re-scaling of axes, was used to assess the compositional gradient-length along gradients of one environmental variable at a time. All ordinations were performed with the computer program CANOCO version 4 (ter Braak & Smilauer, 1998) with rare species down-weighted.

Quantitative transfer functions were developed using weighted averaging partial least squares regression (WA-PLS) and partial least squares regression (PLS). The optimal number of components to include in the transfer functions was assessed by leave-one-out jack-knifing (ter Braak et al., 1993). The best transfer functions were selected as those producing the lowest root mean square error of prediction (RMSEP), but only if the RMSEP was improved by at least 5% with addition of an extra partial least squares component (Birks, 1998). Transfer functions were also selected to have a high coefficient of determination (r^2_{jack}) and low mean and maximum bias. They were developed using the computer program CALIBRATE version 0.81 (ter Braak et al., 1993; Juggins, 1997). Sample specific errors associated with inferences from the transfer functions were calculated using the computer program WA-PLS (version 1.0, Juggins, unpublished program) and 100 random jack-knifed permutations.

Diatom taxon responses were modelled along the gradients of pH and July T in the 100-lake training-set using Gaussian logit regression (GLR). GLR is a form of a generalised linear model that fits a Gaussian-like species response curve to the data (Birks, 1995). The computer program GLR (version 1.1, Juggins, unpublished program) was used to test the statistical significance of relationships between diatom taxa that were present in at least 10% of the training-set lakes and pH/July T, and to estimate the optima and tolerances of each taxon. The significance of each optimum was also tested using a maximum-likelihood model with Poisson error distribution (Huisman et al., 1993) performed by the Huisman-Olff-Fresco (HOF) program (version 2.2, Oksanen, unpublished program). Methods based on HOF and GLR can provide better estimated optima than WA-based methods, because they can assess statistical significance of the optima (Birks, 1995). Optima and tolerances were only reported if both methods exhibited significant fits (for details see Rosén et al., 2000a).

Intra-set cross-validation

A method of intra-set cross-validation was used to assess the performance of diatom-based transfer functions in an independent manner. To do this, the full 100-lake training-set was divided into two data sets. Transfer functions were developed using an 85-lake training-set and applied to diatom assemblages in surface sediment samples from the 15-lake test-set. Diatom-inferred values and their sample specific errors (i.e., RMSEP based on jack-knifing) were then calculated for each of the 15 test lakes and compared with the median values and seasonal variation of lake-water chemistry conditions (pH), or with July T calculated from Climate Normals (1961–1990).

Results

Distribution of diatoms along the altitudinal gradient

Diatom assemblages displayed marked floristic changes along the altitudinal gradient of the 100-lake training-set (Figure 2). Several taxa were restricted mainly to sites above tree-limit (> 650 m a.s.l.), including taxa belonging to *Achnanthes* (e.g., *A. curtissima*, *A. levanderi*, *A. subatomoides*, *A. nodosa*), small *Fragilaria* (e.g., *F. pinnata*, *F. brevistriata*), *Aulacoseira* taxa (*A. distans* var. *alpigena*, *A. distans* var. *nivalis*), *Navicula* (*N. pseudoscutiformis*, *N. digitulus*, *N. minima* var. *muralis*) and *Cymbella* (*C. minuta*, *C. silesiaca*, *C. amphicephala*). Taxa exhibiting highest abundances in lakes within the birch-forest zone (c. 500–650 m a.s.l.) included *Brachysira brebissonii*, *B. vitrea* and *Frustulia rhomboides* var. *saxonica*. However, these taxa were also common in lakes within the coniferous forest (c. < 500 m a.s.l.). *Fragilaria lapponica*, *F. elliptica*, *Navicula hoefleri* (*sensu* Ross & Sims) and *Pinnularia abaujensis* were restricted mainly to lakes within the coniferous forest. Other species (e.g., *Pinnularia biceps*, *Fragilaria pseudoconstruens*, *Achnanthes minutissima* and *Nitzschia perminuta*) were distributed over a broad altitudinal range and were present in almost all of the lakes. Most of the diatom taxa encountered in the training-set lakes were those that are known to occupy mainly benthic habitats and planktonic taxa were relatively rare.

Ordination

Of the 293 taxa encountered in the 100-lake training-set, 157 fulfilled the criteria to be included in the numerical analyses. DCA axes 1 and 2 included floristic gradient lengths of 3.1 and 2.9 SD units, respectively, and indicated that numerical methods based on a unimodal response model were most appropriate (ter Braak & Prentice, 1988).

The forward selection option within CCA identified a set of 5 variables (pH, LOI, July T, lake depth, potassium (K) concentration) that explained the variation in the diatom data almost as well as the complete set of 15 environmental variables included in the statistical analyses. The sum of canonical eigenvalues was 0.545, and compared with DCA of the species data alone, where the sum of all unconstrained eigenvalues was 2.736, the CCA with forward-selected environmental variables explained 20.0% of the variation in the species data.

The CCA ordination with the 5 forward-selected variables illustrated that the composition of diatom assemblages differed markedly between lakes situated above and below the tree-limit (Figure 3a). Diatom assemblages in lakes above the tree-limit were positioned mainly in the upper two quadrants of the biplot, whereas lakes below the tree-limit were positioned in the lower two quadrants. But, CCA was unable to distinguish between diatom assemblages from lakes situated within birch- and coniferous-forests (Figure 3a). Similar to results presented in Figure 2, CCA identified *Achnanthes suchlandtii* (#22), *A. subatomoides* (#21), *Amphora fogediana* (#23), *Cymbella minuta* (#62), *Navicula digitulus* (#105), *N. pseudoscutiformis* (#118), *N. schmassmannii* (#121) and *Pinnularia balfouriana* (#142) as taxa that were restricted mainly to cold-environment lakes above tree-limit (i.e., were positioned in the upper right-hand quadrant, Figure 3b). *Navicula jaagii* (#110), *N. hassiaca* (#107) and *Pinnularia abaujensis* (#141) were most common in lakes with forested catchments and warmer climatic conditions. Overall, relationships between diatoms and catchment vegetation appeared to be closely related to temperature, as demonstrated by the separation of sample scores in Figure 3a along the gradient of July T. Interestingly, relationships between diatoms and climatic conditions were independent of lake-water pH and pH-related variables, as indicated by orthogonal vectors for July T and pH.

A series of partial CCAs, run with one explanatory variable at a time, indicated that pH was the strongest variable and captured 8.0% of the variation in the species data, followed by LOI (5.9%) and July T (4.8%). Interestingly, DOC concentration explained comparatively little variation in diatom assemblage composition (2.9%; Table 2). The ratios between the first (constrained) and the second (unconstrained) eigenvalue were calculated to assess which variables could be used for transfer function development. The ratios for each of these variables were not greatly altered after removing the effects of each of the other variables by entering them as covariables in a series of partial CCAs (Table 2). Diatom assemblages demonstrated significant and independent correlations with each of the environmental variables of interest. Moreover, relationships between diatoms and July T were independent of variables related to physical (LOI and lake depth) or chemical (pH, conductivity, DOC, Si, Ca, Mg, Na, K, SO_4 and Cl) conditions of the aquatic environments. Similarly, relationships between diatoms and physical lake variables (LOI or lake depth) were independent of variables related to temperature (July T, Jan T and Water

Figure 2. Distribution of the 41 most abundant taxa within the training-set. The lakes are ordered according to their elevation.

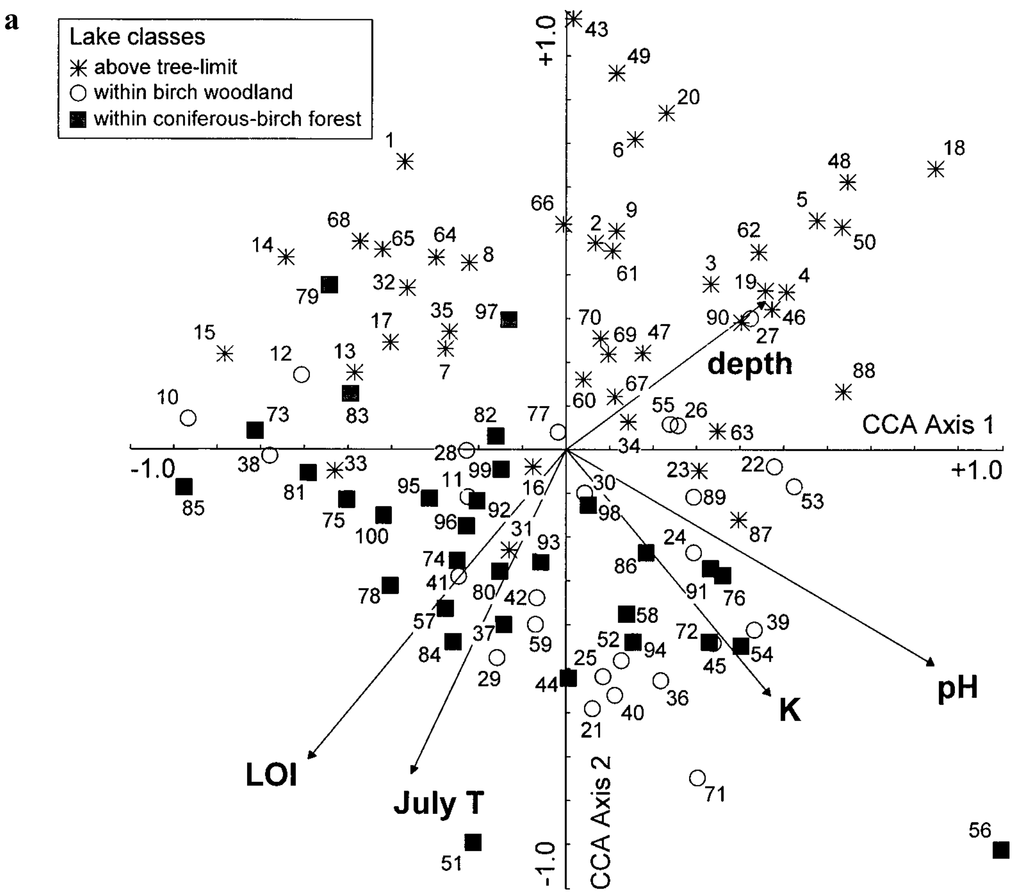

Figure 3. (a) Canonical correspondence analysis (CCA) ordination biplot with site scores and forward selected environmental variables of the 100 lake-set from northern Sweden. Lakes located above present tree-limit are represented as stars, lakes in the mountain birch woodland as circles and lakes in the mixed coniferous-birch forest as squares. Lake numbers refer to lakes indicated in Figure 1 and in Table 1 of Larocque et al. (in press). (b) Canonical correspondence analysis (CCA) ordination biplot with species scores and forward selected environmental variables of the 100 lake-set from northern Sweden. Names of numbered diatom taxa are listed in Table 4.

T) and water chemistry (pH, conductivity, DOC, Si, Ca, Mg, Na, K, SO_4 and Cl; Table 2).

Transfer functions

Transfer functions were developed for pH and July T, because they showed relatively high ratio's (λ_1/λ_2) of the first constrained to the second unconstrained axes in CCA (Table 2), and they were the variables of greatest paleolimnological and paleoclimatic interest.

The gradient length in the diatom data was 2.84 SD units for pH and 1.41 SD units for July T, as assessed by DCCA. Transfer functions for pH were based only on WA-PLS, because the floristic gradient length was > 2 SD units. The gradient length for July T was less than 2 SD units, suggesting a linear relationship between diatoms and July T (ter Braak et al., 1993; ter Braak, 1995). However, we used both linear and unimodal methods (i.e., PLS and WA-PLS) when developing the diatom–July T transfer function, because it has been shown that WA-PLS usually provides a better transfer function than PLS even when floristic gradients are < 2 SD units (Birks, 1995; ter Braak, 1995).

One lake (Alanen Laanijärvi, #56) was identified as an outlier with respect to pH and was removed for the pH calibration due to its exceptionally high pH (> 3 SD units from the mean pH of the 100-lake data set) and its high leverage. No obvious outlier was identified for July T, and consequently all 100 lakes were used to develop transfer functions for this variable.

For both variables (pH, July T) the best transfer function or the 'minimum adequate model' (Birks, 1998) was a two-component WA-PLS model. Significant relationships could be established between diatom assem-

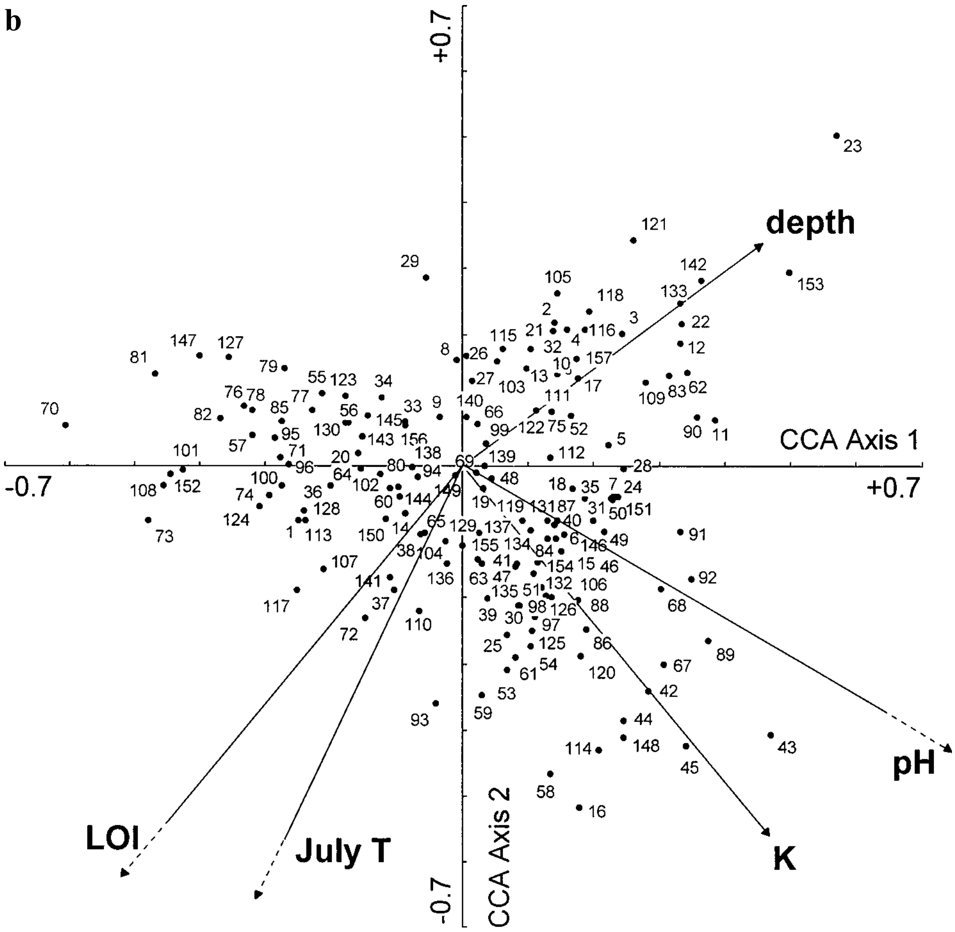

Figure 3. Continued.

blages and pH ($r^2_{\text{jack-knifed}} = 0.77$, RMSEP = 0.19), and July T ($r^2_{\text{jack-knifed}} = 0.75$, RMSEP = 0.96 °C; Table 3, Figure 4). The two-component WA-PLS models also provided the lowest mean and maximum biases, and for July T the WA-PLS model performed slightly better than the PLS model.

Diatom optima for pH and July T are listed in Table 4. Of the 119 diatom taxa that occurred in at least 10 lakes in the training set, 98 had significant optima for pH (= 82.4%) and 90 had significant optima for July T (= 76.5%).

Diatom taxa with the highest significant optima for pH (> 7.46) included *Achnanthes minutissima*, *A. pusilla*, *Brachysira zellensis*, *Cyclotella rossii*, *Cymbella minuta*, *Navicula radiosa*, *Nitzschia dissipata* var. *dissipata*, and *N. palea*. Species with significant pH-optima below the sampled pH gradient (< 5.79 units) included *Achnanthes altaica*, *Eunotia hexaglyphis*, *E. pectinalis* var. *minor*, *E. rhomboidea*, *E. vanheurckii*, *Navicula hoefleri* (*sensu* Ross & Sims), *N. mediocris*, *Neidium affine* var. *longiceps*, *Nitzschia perminuta* and *Tabellaria flocculosa* var. *flocculosa*. Overall, 11 taxa showed statistically significant skewed unimodal responses to pH, 34 taxa had significant symmetrical unimodal responses, 53 taxa had significant sigmoidal increasing or decreasing responses with optima beyond the sampled pH gradient, and 21 taxa showed no significant relationships with pH.

Diatom taxa with significant optima were more common in lakes with warm July T (> 14.7 °C) and included *Aulacoseira perglabra*, *Eunotia curvata*, *Fragilaria lapponica*, *Navicula hoefleri* (*sensu* Ross & Sims), *Navicula vitiosa* and *Pinnularia nodosa*. The cold-water indicators (i.e., significant optima < 7.0 °C) were *Achnanthes curtissima*, *A. laterostrata*, *A. suchlandtii*, *Aulacoseria distans* var. *nivalis*, *Gomphonema parvulum*, *Navicula indifferens* and *N. schmassmannii*. Overall, 13 taxa had statistically significant skewed unimodal responses to

Table 2. Summary of partial canonical correspondence analysis (partial CCA) of surface diatom data (n = 100) to assess the independence of selected environmental variables

Variable	Covariable	Eigenvalue axis 1 (λ_1)	Eigenvalue axis 2 (λ_2)	Ratio (λ_1/λ_2)	% Variation explained on axis 1	p-value (99 permutations)
pH	–	0.220	0.272	0.809	8.0	0.01(*)
pH	LOI	0.205	0.182	1.126	8.0	0.01(*)
pH	July T	0.218	0.239	0.912	8.4	0.01(*)
pH	DOC	0.220	0.245	0.898	8.3	0.01(*)
pH	phys	0.198	0.168	1.179	7.9	0.01(*)
pH	temp	0.223	0.195	1.144	8.9	0.01(*)
pH	phys, temp	0.190	0.153	1.242	8.1	0.01(*)
LOI	–	0.162	0.286	0.566	5.9	0.01(*)
LOI	pH	0.147	0.182	0.807	5.8	0.01(*)
LOI	July T	0.110	0.279	0.394	4.2	0.01(*)
LOI	DOC	0.125	0.281	0.445	4.7	0.01(*)
LOI	temp	0.089	0.274	0.325	3.5	0.01(*)
LOI	chem	0.077	0.159	0.484	3.7	0.01(*)
LOI	temp, chem	0.056	0.139	0.403	2.8	0.01(*)
July T	–	0.133	0.333	0.399	4.8	0.01(*)
July T	pH	0.131	0.239	0.548	5.2	0.01(*)
July T	LOI	0.081	0.279	0.290	3.2	0.01(*)
July T	DOC	0.080	0.333	0.240	3.0	0.01(*)
July T	phys	0.075	0.270	0.278	3.0	0.01(*)
July T	chem	0.057	0.199	0.286	2.7	0.01(*)
July T	phys, chem	0.038	0.140	0.271	1.9	0.02(*)
DOC	–	0.080	0.342	0.234	2.9	0.01(*)
DOC	pH	0.080	0.245	0.327	3.2	0.01(*)
DOC	LOI	0.044	0.281	0.157	1.7	0.02(*)
DOC	July T	0.027	0.333	0.081	1.0	0.42
DOC	phys	0.051	0.265	0.192	2.0	0.01(*)
DOC	temp	0.028	0.321	0.087	1.1	0.39
DOC	phys, temp	0.029	0.260	0.111	1.2	0.25

Each variable is used as the sole constraining variable at a time and with included covariables. The first constrained eigenvalue (λ_1), the second unconstrained eigenvalue (λ_2), their ratio (λ_1/λ_2), percentage variation explained by each environmental variable, and results of Monte Carlo permutation tests (99 unrestricted permutations) are reported. *indicates significant relationships (p-value < 0.05). The variable 'phys' includes the physical variables depth and LOI; 'chem' includes water chemistry variables pH, DOC, Si, cond, Ca, Mg, Na, K, Sulphate, and Cl; and 'temp' includes Water T, July T, and Jan T.

July T, 41 had significant symmetrical unimodal responses, 36 had significant sigmoidal increasing or decreasing responses with optima beyond the sampled temperature gradient, and 29 taxa showed no statistically significant relationships to temperature.

Intra-set cross-validation of the pH and temperature transfer functions

The 85-lake training-set included similar ranges and distributions of physical, chemical and climatic variables (Table 5) as the full 100-lake data set (Table 1), and transfer functions were based on the same 157 diatom taxa and used the same two-component WA-PLS methods. Additionally, the 15-lake test-set included almost as broad a range of lake-water pH and climatic conditions as the full 100-lake set, and so allowed us to assess transfer function performance over the full range of present-day conditions (Tables 1 & 5).

Performance statistics of the 85-lake pH transfer function were almost identical to those of the 100-lake training-set ($r^2_{\text{jack-knifed}}$ = 0.78 vs. 0.77, RMSEP = 0.19 vs. 0.19, mean bias = 0.013 vs. 0.009, max. bias = 0.28 vs. 0.31, for 85- vs. 100-lake transfer functions, respectively). Similarly, the 85-lake July T transfer function produced statistical performance that closely approximated that of the 100-lake set ($r^2_{\text{jack-knifed}}$ = 0.70 vs. 0.75, RMSEP = 1.03 vs. 0.96 °C, mean bias = 0.060 vs.

Table 3. Performance of apparent and cross-validated statistics of the transfer functions for lake-water pH based on 99 lakes and mean July air temperature (July T) based on 100 lakes

Variable	pH	July T
Model type	WA-PLS	WA-PLS
# of PLS components	2	2
Apparent		
r^2_{apparent}	0.87	0.86
RMSE	0.15	0.67°C
RMSE in % of gradient length	9.0%	8.7%
Avg. bias	0.004	0.024°C
Max. bias	0.21	0.49°C
Cross-validation jack-knifing		
$r^2_{\text{jack-knifed}}$	0.77	0.75
RMSEP	0.19	0.96°C
RMSEP in % of gradient length	11.4%	12.4%
Avg. bias	0.009	0.042°C
Max. bias	0.30	1.37°C

0.042 °C, max. bias = 1.51 vs. 1.37 °C, respectively).

Overall, intra-set cross-validation demonstrated that the pH and temperature transfer functions based on the 85-lake training set provided accurate and robust estimates of measured values in the 15 test-set lakes (Figures 5 & 6). For example, diatom-inferred pH values, including prediction errors, fell within 1 SD unit of the range of inter-annual variability in all but one of the test lakes (#69; Figure 5). Additionally, diatom-inferred pH values corresponded closely with the median measured values. The minimum difference between diatom-inferred and measured pH in the 15-lake test-set was 0.01 pH units, and the maximum discrepancy was 0.54 pH units. On average, the differences between diatom-inferred and median measured pH in the 15 test lakes was only 0.16 of a pH unit (SD = 0.14), which was less than the inter-annual variability in pH in each of the 15 test-set lakes (as estimated by one standard deviation unit; Figure 5).

The 85-lake transfer function for July T performed similarly well in cross-validation as the transfer functions for pH. In all but two of the test lakes (#51, #59), the temperatures derived from the weather stations fell within the RMSEP of the diatom-inferred temperatures (Figure 6). In 11 of the 15 test-set lakes, diatom-inferred temperatures corresponded extremely closely to the values derived from weather stations. In general, diatom-temperature inferences from cold (high elevation) sites appeared to be more reliable than those from warm (low elevation) sites. Overall, the minimum discrepancy between the diatom-inferred and actual temperature for the 15-lake test-set was 0.02 °C and the maximum discrepancy was 2.24 °C. On average, the difference was 0.62 °C (SD = 0.55 °C).

Discussion

The need for accurate long-term climatic records has spurred the proliferation of paleolimnological temperature transfer functions (e.g., Pienitz et al., 1995; Lotter et al., 1997; Weckström et al., 1997; Olander et al., 1999; Brooks & Birks, 2000; Rosén et al., 2000a; Larocque et al., in press; Bigler et al., 2000). In most cases, these transfer functions show considerable potential as useful tools to reconstruct past changes in July T, but there is a need to critically assess them. To date, however, there have been few critical assessments, in part because it is extremely difficult to experimentally determine direct relationships between diatoms and temperature (De Nicola, 1996), and because of logistical constraints associated with the remote setting of existing training-sets (Lotter et al., 1999).

Overall, our analysis of diatom assemblages in the surface sediments from lakes across steep ecological and climatic gradients in Swedish Lapland indicates that diatoms can provide useful and reliable quantitative estimates of past changes in July T, as assessed by three main lines of evidence. First, CCA demonstrated that diatom-temperature relationships were statistically significant and independent of physical and chemical conditions of the lakes. We realise, however, that diatom communities are not responding directly to air temperature, but to temperature-related gradients in thermal and ecological conditions of lakes. Logically, aquatic biota respond to their thermal environment, and the summer heat budget of a lake is strongly determined by air temperature (Livingstone & Lotter, 1998). In Northern Scandinavia, air temperature is strongly correlated with many climate-related variables including water temperature, summer heat budget, altitude, and duration of the ice-free season (Korhola et al., 2000; Larocque et al., in press). One of the basic requirements for transfer functions is that the environmental variable to be reconstructed is an ecologically important variable in the system of interest (Birks et al., 1990). We therefore conclude that it is statistically and ecologically reasonable to quantify July T from diatoms, because direct gradient analyses demonstrated that diatom distributions are correlated with July T and that the correlations are independent of other strong limnological variables as pH or DOC.

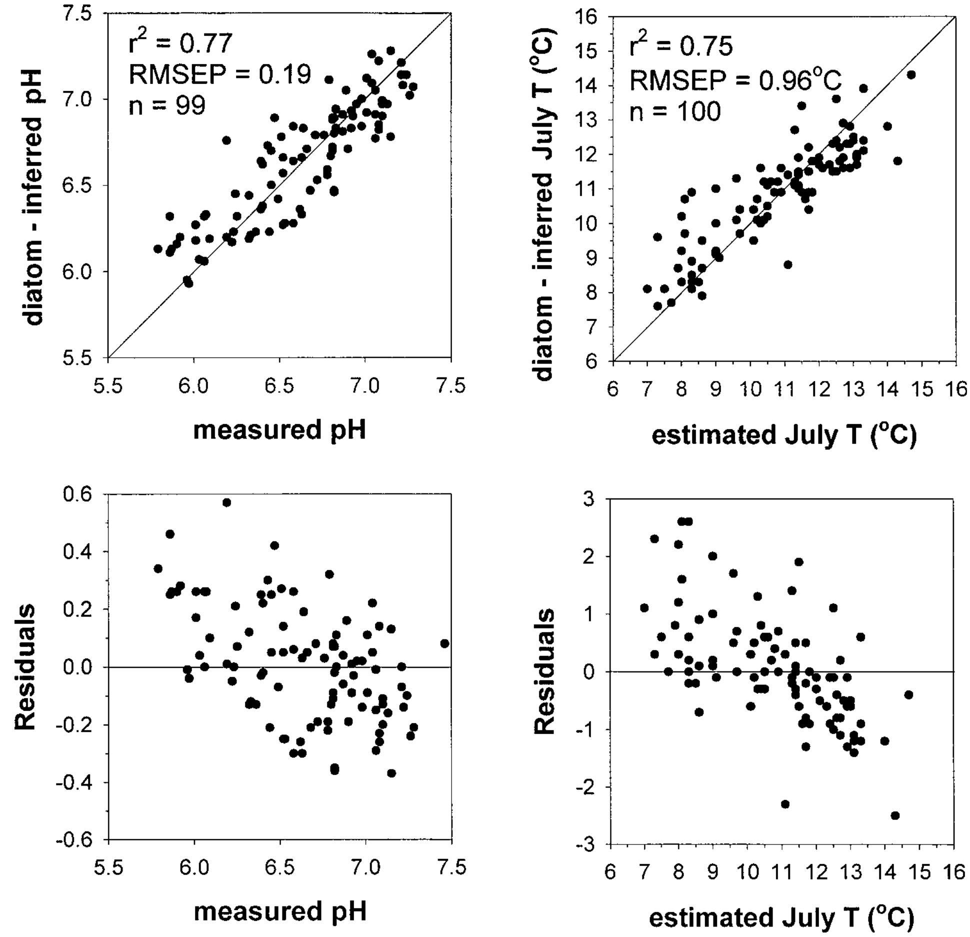

Figure 4. Relationships between measured and diatom-inferred lake water pH and between estimated mean July air temperatures (July T) and diatom-inferred July T using 2-component weighted-averaging partial least squares (WA-PLS) regression and calibration models.

Second, performance statistics (r^2_{jack}, RMSEP as % of gradient length) of the diatom-transfer function for July T were comparable to those of the diatom-pH transfer function, based on the complete training-set (r^2_{jack} = 0.75 vs. 0.77; RMSEP as % of gradient = 12.5 vs. 11.4, respectively). Moreover, the broadly similar number of diatom taxa with significant optima for July T and lake-water pH (90 vs. 98 taxa, respectively; as assessed by HOF and GLR) suggests that diatom-temperature relationships may be about as strong as diatom-pH relationships. Because diatom-pH relationships and pH transfer functions are widely considered to be strong, accurate and reliable (e.g., Stevenson et al., 1989; Renberg et al., 1993; Hall & Smol 1996; Korhola et al., 1999), our findings suggest that diatom-temperature transfer functions may provide a comparably sensitive and robust reconstruction tool.

Third, intra-set cross-validation demonstrated that the July T transfer function based on 85 training-lakes accurately estimated actual temperatures (derived from weather stations) in an independent set of 15 test-lakes. Diatom-inferred temperatures were relatively close (within 0.6 °C) to values based on climate stations in 11 of the 15 test-lakes, and the temperatures fell outside the sample specific errors of diatom inferences in only 2 of the lakes (i.e., lakes #51, 59). Interestingly, the transfer function performed slightly better in colder, high elevation sites above tree-limit (< 10 °C) than in warmer forested sites. In fact, the two lakes with poorest correspondence between actual and diatom-inferred temperature in cross-validation (lakes #51, 59) were from the warm end of the temperature gradient. The reason for this remains uncertain, but one possibility is that correlations between air and lakewater temperatures are greater in colder, unforested catchments than in warmer, forested catchments (Livingstone, per-

Table 4. Estimated optima and tolerances for diatom taxa occurring in at least 10 lakes using Gaussian logit regression (GLR) and Huisman-Olff-Fresco (HOF) models for pH (n = 99) and estimated mean July air temperature (n = 100)

Nr.	*Taxon*	Authority	# occ	pH (units)		July T (°C)	
				Opt.	Tol.	Opt.	Tol.
1	*Achnanthes altaica*	(Poretzky) Cleve-Euler 1953	31	<5.79		12.4	1.8
2	*Achnanthes austriaca* var. *helvetica*	Hustedt 1933	29	–		–	
3	*Achnanthes carissima*	Lange-Bertalot in Lange-Bertalot & Krammer 1989	16	6.89	0.26	–	
4	*Achnanthes curtissima*	Carter 1963	64	6.63	0.37	<7.0	
5	*Achnanthes didyma*	Hustedt 1933	30	7.10	0.41	–	
6	*Achnanthes flexella*	(Kützing) Brun 1880	22	7.13	0.39	–	
7	*Achnanthes impexiformis*	Lange-Bertalot in Lange-Bertalot & Krammer 1989	13	6.88	0.27	–	
8	*Achnanthes kriegeri*	Krasske 1943	16	–		–	
9	*Achnanthes kuelbsii*	Lange-Bertalot in Lange-Bertalot & Krammer 1989	15	6.61	0.36	10.0	1.1
12	*Achnanthes laterostrata*	Hustedt 1933	24	6.89	0.21	<7.0	
13	*Achnanthes levanderi*	Hustedt 1933	80	–		7.1	2.7
14	*Achnanthes marginulata*	Grunow in Cleve & Grunow 1880	40	–		11.0	1.8
15	*Achnanthes minutissima*	Kützing 1833	80	>7.46		–	
16	*Achnanthes minutissima* var. *macrocephala*	Hustedt 1937	11	–		12.0	0.8
17	*Achnanthes nodosa*	Cleve 1900	37	6.76	0.20	9.2	1.6
18	*Achnanthes pusilla*	(Grunow) De Toni 1891	69	>7.46		9.4	2.4
19	*Achnanthes saccula*	Carter in Carter & Bailey-Watts 1981	18	6.62	0.40	11.0	1.4
20	*Achnanthes scotica*	Flower 1990	24	6.42	0.30	10.4	0.9
21	*Achnanthes subatomoides*	(Hustedt) Lange-Bertalot & Archibald in Krammer & Lange-Bertalot 1985	68	–		–	
22	*Achnanthes suchlandtii*	Hustedt 1933	23	7.11	0.51	<7.0	
24	*Amphora libyca*	Ehrenberg 1840	32	6.79	0.25	–	
26	*Aulacoseira distans* var. *alpigena*	(Grunow in Van Heurck) Simonsen 1979	52	6.46	0.40	–	
27	*Aulacoseira distans* var. *distans*	(Ehrenberg) Simonsen 1979	26	–		–	
28	*Aulacoseira distans* var. *nivalis*	(Smith) Haworth 1988	48	6.12	0.46	<7.0	
31	*Aulacoseira italica* var. *valida*	(Grunow) Simonsen 1979	11	7.08	0.45	12.1	1.5
32	*Aulacoseira lirata*	(Ehrenberg) Ross in Hartley 1986	21	6.70	0.32	9.8	1.6
34	*Aulacoseira perglabra*	(Oestrup) Haworth 1988	30	–		>14.7	
35	*Aulacoseira subarctica* type 2	SWAP	15	6.88	0.38	12.0	1.4
36	*Brachysira brebissonii*	Ross in Hartley 1986	68	6.30	0.44	10.8	1.6
37	*Brachysira styriaca*	(Grunow in Van Heurck) Ross in Hartley 1986	24	–		11.5	1.2
38	*Brachysira vitrea*	(Grunow) Ross in Hartley 1986	81	–		11.2	2.0
39	*Brachysira zellensis*	(Grunow) Round & Mann 1981	14	>7.46		11.8	1.1
40	*Caloneis bacillum*	(Grunow) Cleve 1894	10	7.27	0.43	10.6	0.9
41	*Caloneis tenuis*	(Gregory) Krammer 1985	11	–		11.0	1.1
42	*Cyclotella antiqua*	Smith 1853	11	7.27	0.15	11.3	0.8
44	*Cyclotella comta*	Grunow in Van Heurck 1882	16	7.32	0.37	13.5	1.5
46	*Cyclotella glomerata*	Bachmann 1911	18	6.91	0.22	–	
49	*Cyclotella rossii*	Håkansson 1990	15	>7.46		11.6	0.9
52	*Cymbella amphicephala*	Naegeli in Kützing 1849	24	7.05	0.41	9.6	1.8
53	*Cymbella cesatii*	(Rabenhorst) Grunow 1881	29	–		11.8	1.0
54	*Cymbella descripta*	(Hustedt) Krammer & Lange-Bertalot 1985	41	7.26	0.39	11.6	1.4
56	*Cymbella gaeumannii*	Meister 1934	63	6.21	0.44	10.3	1.6
57	*Cymbella hebridica*	(Grunow in Cleve) Cleve 1894	36	6.03	0.40	11.0	2.0
58	*Cymbella incerta*	(Grunow) Cleve 1894	14	–		12.0	0.8
60	*Cymbella lunata*	W. Smith in Greve 1855	83	6.45	0.49	11.1	2.1
61	*Cymbella microcephala*	Grunow in Van Heurck 1880	35	–		11.8	1.3
62	*Cymbella minuta*	Hilse ex Rabenhorst 1862	27	>7.46		–	

Table 4. Continued

Nr.	*Taxon*	Authority	# occ	pH (units)		July T (°C)	
				Opt.	Tol.	Opt.	Tol.
63	*Cymbella naviculiformis*	(Auerswald) Cleve 1894	13	–		–	
64	*Cymbella perpusilla*	Cleve 1895	34	6.16	0.50	14.2	4.0
66	*Cymbella silesiaca*	Bleisch in Rabenhorst 1864	50	–		–	
67	*Denticula kuetzingii*	Grunow 1862	17	7.42	0.31	11.5	1.7
71	*Eunotia curvata*	(Kützing) Lagerstedt 1884	34	–		>14.7	
74	*Eunotia exigua*	(Brébisson ex. Kützing) Rabenhorst 1864	41	–		11.9	1.7
75	*Eunotia flexuosa*	(Brébisson) Kützing 1849	16	6.75	0.25	9.3	1.7
76	*Eunotia hexaglyphis*	Ehrenberg 1854	10	<5.79		10.9	1.3
77	*Eunotia incisa*	Gregory 1854	21	–		11.6	1.8
78	*Eunotia naegelii*	Migula in Thomé 1907	22	–		10.6	1.4
79	*Eunotia pectinalis* var. *minor*	(Kützing) Rabenhorst 1864	32	<5.79		9.6	1.6
80	*Eunotia praerupta*	Ehrenberg 1843 sensu lato	32	–		–	
81	*Eunotia rhomboidea*	Hustedt 1950	17	<5.79		–	
82	*Eunotia vanheurckii*	Patrick 1958	35	<5.79		10.6	1.7
83	*Fragilaria brevistriata*	Grunow in Van Heurck 1885	39	7.13	0.64	–	
84	*Fragilaria capucina*	Desmaziéres 1925	20	7.13	0.42	–	
87	*Fragilaria construens* f. *venter*	(Ehrenberg) Hustedt 1957	31	–		–	
88	*Fragilaria elliptica*	Schumann 1867	22	–		–	
89	*Fragilaria lapponica*	Grunow in Van Heurck 1881	10	7.34	0.31	>14.7	
90	*Fragilaria pinnata*	Ehrenberg 1843	40	6.99	0.34	–	
91	*Fragilaria pseudoconstruens*	Marciniak 1982	45	7.31	0.27	–	
94	*Fragilaria virescens* var. *exigua*	Grunow in Van Heurck 1881	67	6.38	0.52	11.4	2.0
95	*Frustulia rhomboides*	(Ehrenberg) De Toni 1891	39	6.01	0.48	11.3	1.7
96	*Frustulia rhomboides* var. *saxonica*	(Rabenhorst) De Toni 1891	73	5.82	0.57	11.0	1.6
97	*Gomphonema acuminatum* var. *coronatum*	(Ehrenberg) Smith 1853	20	7.33	0.44	–	
98	*Gomphonema gracile*	Ehrenberg 1838	12	6.96	0.17	–	
99	*Gomphonema parvulum*	(Kützing) Kützing 1849	36	6.78	0.44	<7.0	5.3
100	*Navicula begerii*	(Krasske) Krasske 1932	12	6.35	0.37	11.4	1.3
101	*Navicula bremensis*	Hustedt 1957	27	5.90	0.46	11.5	1.5
102	*Navicula bryophila*	Petersen 1928	64	–		–	
103	*Navicula cocconeiformis*	Gregory ex Greville 1856	14	–		–	
104	*Navicula cryptocephala*	Kützing 1844	64	6.54	0.43	–	
105	*Navicula digitulus*	Hustedt 1943	45	6.61	0.36	7.6	2.2
107	*Navicula hassiaca*	Krasske 1925	16	–		12.7	0.6
108	*Navicula hoeflerii* (*sensu* Ross & Sims)	Ross & Sims 1978	28	<5.79		>14.7	
109	*Navicula indifferens*	Hustedt 1942	22	7.01	0.39	<7.0	
110	*Navicula jaagii*	Meister 1934	14	–		–	
112	*Navicula laevissima*	Kützing 1844	39	6.73	0.33	–	
113	*Navicula mediocris*	Krasske 1932	56	<5.79		12.5	1.7
115	*Navicula minima*	Grunow in Van Heurck 1880	17	–		–	
116	*Navicula minuscula* var. *muralis*	(Grunow) Lange-Bertalot 1981	26	–		–	
118	*Navicula pseudoscutiformis*	Hustedt 1930	36	6.67	0.32	–	
119	*Navicula pupula*	Kützing 1844	53	6.82	0.35	–	
120	*Navicula radiosa*	Kützing 1844	40	>7.46		11.5	1.9
121	*Navicula schmassmannii*	Hustedt 1943	25	–		<7.0	
122	*Navicula seminulum*	Grunow 1860	52	6.71	0.42	–	
123	*Navicula seminulum* var. *intermedia*	Hustedt 1942	33	5.93	0.43	–	
124	*Navicula subtilissima*	Cleve 1891	49	–		11.7	1.5
126	*Navicula vitiosa*	Schimanski 1978	16	7.15	0.52	>14.7	
128	*Neidium affine* var. *longiceps*	(Gregory) Cleve 1894	20	<5.79		–	
129	*Neidium alpinum*	Hustedt 1943	11	–		10.8	0.9
130	*Neidium ampliatum*	(Ehrenberg) Krammer 1985	30	5.90	0.69	10.1	1.7
131	*Nitzschia alpina*	Hustedt 1943	33	7.39	0.58	10.4	1.6
132	*Nitzschia angustata*	(Smith) Grunow in Cleve & Grunow 1880	17	6.86	0.25	–	

Table 4. Continued

Nr.	*Taxon*	Authority	# occ	pH (units)		July T (°C)	
				Opt.	Tol.	Opt.	Tol.
134	*Nitzschia dissipata* var. *dissipata*	(Kützing) Grunow 1862	37	>7.46		–	
135	*Nitzschia fonticola*	Grunow in Cleve & Möller 1879	59	–		11.0	1.4
136	*Nitzschia gracilis*	Hantzsch 1860	45	–		–	
137	*Nitzschia palea*	(Kützing) Smith 1856	28	>7.46		10.6	1.1
138	*Nitzschia perminuta*	(Grunow) Peragallo 1903	87	<5.79		–	
141	*Pinnularia abaujensis*	(Pantocsek) Ross 1947	36	–		14.0	2.1
142	*Pinnularia balfouriana*	Grunow ex Cleve 1895	12	6.89	0.24	8.8	1.0
143	*Pinnularia biceps*	Gregory 1856	87	5.88	0.64	–	
144	*Pinnularia microstauron* var. *brebissonii*	(Kützing) Mayer 1912	69	–		12.5	3.6
145	*Pinnularia microstauron* var. *microstauron*	(Ehrenberg) Cleve 1891	23	6.13	0.55	–	
146	*Pinnularia nodosa*	(Ehrenberg) Smith 1856	15	6.79	0.24	>14.7	
147	*Pinnularia viridis*	(Nitzsch) Ehrenberg 1843	15	5.79	0.40	10.6	1.3
149	*Stauroneis anceps*	Ehrenberg 1843	62	6.58	0.39	–	
150	*Stauroneis phoenicenteron*	(Nitzsch) Ehrenberg 1843	56	–		–	
154	*Synedra tenera*	Smith 1856	35	6.80	0.29	–	
155	*Tabellaria flocculosa* agg.	SWAP	46	6.90	0.52	–	
156	*Tabellaria flocculosa* var. *flocculosa*	(Roth) Kützing 1844	87	<5.79		10.5	2.2
157	*Tetracyclus lacustris*	Ralfs 1843	20	6.65	0.27	–	

Missing values indicate taxa that show no significant relationship to pH or temperature by GLR and/or HOF. Taxa with estimated optima outside the range of the training set are indicated by > or < signs. Species numbers correspond with Figure 3b

sonal communication). Overall, we conclude that diatom-based transfer functions are able to accurately infer July T from lake sediment samples in Swedish Lapland.

Diatom-temperature relationships observed in the 100-lake training set from Swedish Lapland agree reasonably well with those observed in a similar study by Rosén et al. (2000a), located approximately 200 km to the south. Rosén et al.'s study is the only one that has calculated diatom optima and tolerances for July T using the same underlying statistical approach (GLR/HOF, see Methods section), and offers a rare opportunity to assess if diatom-temperature optima are robust among training-sets. Overall, agreement between optima of individual taxa is good, especially among most species belonging to *Achnanthes*, *Cymbella*, *Fragilaria*, *Eunotia*, and *Tabellaria*. For example, taxa with optima indicating preference for relatively warm temperatures in both training-sets are *Brachysira* species, *Navicula hoefleri*, *Navicula mediocris* and *Navicula radiosa*. Cold temperature indicators for both training-sets included mainly small *Achnanthes* species. However, the temperature optimum for *Aulacoseira perglabra* disagrees between the training-sets, probably due to different taxonomical concepts. In our study, the temperature optimum for *A. perglabra* was > 14.7 °C, whereas Rosén et al. (2000a) suggest it is < 7.0 °C. The pH-optima and tolerances are also very similar among the two training-sets, with *Cyclotella* and *Fragilaria* indicating in general high pH, *Eunotia* and *Tabellaria* species indicating low pH in both training-sets. Overall, these findings suggest that diatom optima for July T and lake-water pH can be accurately quantified using the numerical methods outlined in this paper and that the relationships are largely repeatable among training sets that possess similar ecological and climatic gradients but from different geographic locations.

As identified by Birks (1998), transfer functions (and the training sets from which they are developed) rarely take into account the quality, reliability and representativeness of the environmental data upon which they are based. A central question that has thus far remained relatively unanswered is whether diatom transfer functions that are based on single water-chemistry measurements can accurately represent the central tendency and variability of important limnological variables. We attempted to address this issue using intra-set cross-validation of the pH transfer function, where the 85-lake training-set was based on single-point measurements and the 15-lake test-set included variability in pH over two ice-free seasons. Overall, our analyses demonstrated that diatom-inferred pH closely corresponded with median lake-water pH in 14 out of 15 test-lakes. Moreover, prediction errors associated with diatom-pH inferences were usually within the range of inter-annual variability. These findings indicate that the pH

Table 5. Summary of the environmental characteristics of the 85 lake training-set (a) and the 15 lake test-set (b), as used in intra-set cross-validation

	Variable	Min.	Max.	Mean	Median	S.D.
(a) 85-lake training-set						
Water T	Measured water temp. (°C)	2.4	14.5	10.4	10.5	2.4
July T	Est. mean July air temp. (°C)	7.0	14.7	10.9	11.3	1.9
Jan T	Est. mean January air temp. (°C)	–16.4	–12.9	–14.6	–14.5	0.8
Depth	Max. lake depth (m)	1.5	16.5	5.9	5.0	3.2
LOI	Loss On Ignition (%)	2.9	74.2	42.5	42.4	16.0
pH	pH	5.79	7.46	6.62	6.66	0.42
DOC	DOC (mg/L)	0.7	13.4	3.5	2.8	2.4
Si	Si (mg/L)	0.010	3.060	0.753	0.660	0.727
Cond	Conductivity (μS/m)	0.560	6.440	1.974	1.620	1.292
Ca	Ca (meq/L)	0.010	0.406	0.089	0.064	0.081
Mg	Mg (meq/L)	0.007	0.169	0.039	0.027	0.036
Na	Na (meq/L)	0.011	0.107	0.036	0.032	0.016
K	K (meq/L)	0.003	0.033	0.007	0.006	0.006
Sulphate	SO_4 (meq/L)	0.010	0.361	0.042	0.026	0.055
Chloride	Cl (meq/L)	0.008	0.104	0.026	0.022	0.016
Lat.	Latitude	67.07	68.48	68.10	68.20	0.33
Long.	Longitude	17.67	23.52	19.83	19.57	1.54
Elev.	Elevation (m a.s.l.)	169	1183	585	490	257
(b) 15-lake test-set						
Water T	Measured water temp. (°C)	7.0	17.5	10.3	9.5	3.0
July T	Est. mean July air temp. (°C)	7.5	13.3	10.5	10.1	1.9
Jan T	Est. mean January air temp. (°C)	–16.4	–12.9	–14.8	–15.0	0.9
Depth	Max. lake depth (m)	3.4	15.8	7.4	7.3	3.4
LOI	Loss On Ignition (%)	7.8	87.3	40.5	37.3	22.4
pH	pH	5.96	8.07	6.85	6.83	0.46
DOC	DOC (mg/L)	0.2	8.4	3.4	3.1	2.4
Si	Si (mg/L)	0.710	1.560	0.992	0.920	0.290
Cond	Conductivity (μS/m)	0.640	12.800	2.677	1.980	2.979
Ca	Ca (meq/L)	0.011	1.030	0.155	0.076	0.250
Mg	Mg (meq/L)	0.007	0.175	0.044	0.030	0.043
Na	Na (meq/L)	0.014	0.061	0.034	0.031	0.013
K	K (meq/L)	0.002	0.050	0.011	0.006	0.012
Sulphate	SO_4 (meq/L)	0.009	0.089	0.036	0.031	0.024
Chloride	Cl (meq/L)	0.010	0.042	0.020	0.017	0.008
Lat.	Latitude	67.57	68.45	68.10	68.20	0.28
Long.	Longitude	18.62	21.52	19.73	19.62	0.91
Elev.	Elevation (m a.s.l.)	270	1115	654	710	279

transfer function based on single point-in-time measurements is able to accurately estimate the central tendency of lake-water pH (mean discrepancy = 0.16 pH units). The prediction errors are less than the magnitude of inter-annual variability, even over the relatively broad gradient included in the test set (pH 5.9–7.3).

Overall, the transfer functions for temperature and pH seemed to be biased slightly over their gradient: the models generally overestimate measured values at low pH and temperature and underestimate measured values at high pH and temperature. At least two factors may account for this phenomenon. First, inverse deshrinking is known to introduce such bias between measured and inferred values and, second, bias may be introduced by the statistically relatively short gradients that were sampled (Birks, 1998). However, when compared to other similar training-sets (Rosén et al., 2000a: 120, Table 6), our training-set has one of the lowest prediction errors (RMSEP) and one of the highest coefficients of determination (r^2).

Our future research will apply the transfer functions developed in this study for environmental reconstructions during the Holocene in the Abisko area of northern Sweden. Although our study shows that diatoms potentially provide a relatively reliable method to quantify past temperature changes, many other variables can

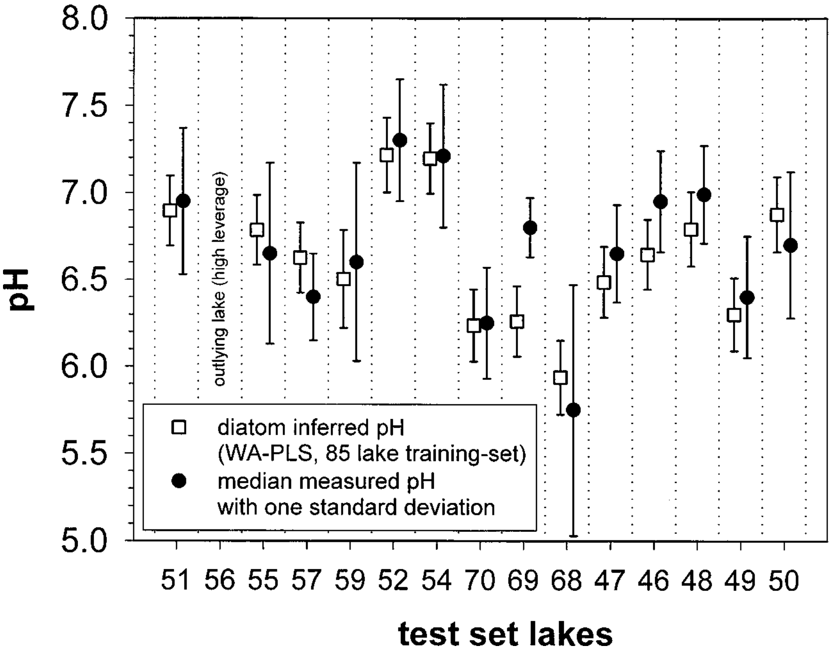

Figure 5. Diatom-based WA-PLS predictions of pH (based on the 85 lake training-set) for 15 independent test lakes with lake specific errors (root mean square errors of prediction based on jack-knifing, indicated as error bars) compared to the measured pH values for two seasons with one standard deviation unit.

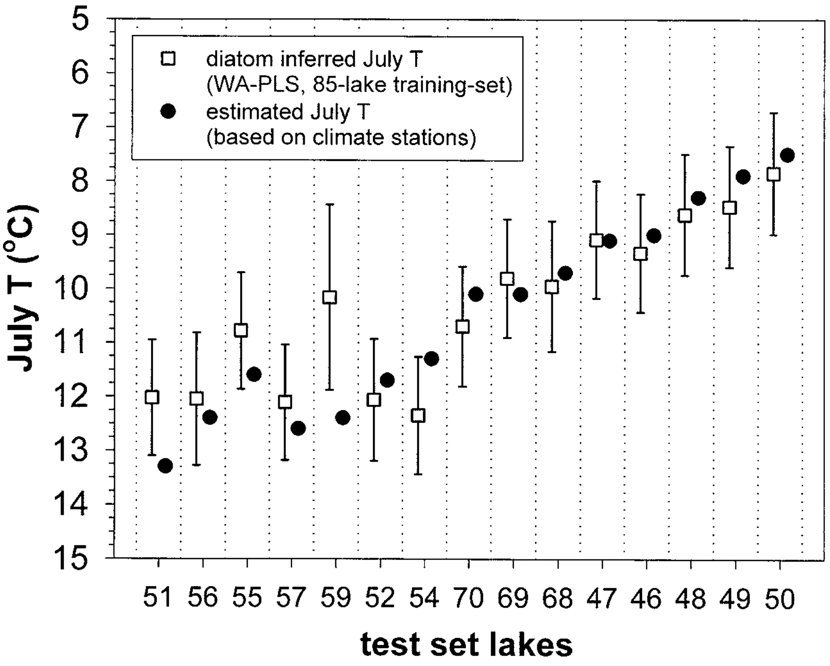

Figure 6. Diatom-based WA-PLS predictions of air temperature (based on 85 lake training-set) for 15 independent test lakes with lake specific errors (indicated as error bars) compared to the estimated mean July air temperature (July T) based on the last normal period (1961–1990) record of the nearest situated climate stations and a lapse rate of 0.57 °C/100 m.

strongly influence assemblage composition. We, therefore, continue to suggest that diatom-temperature transfer functions should be combined with other proxy methods.

Acknowledgements

The funds were provided through the Climate Impacts Research Centre (CIRC) within Miljö och Rymdinstitutet (MRI). We are grateful to A. Odland for the estimations of air temperature at all our training-set lakes, to H. J. B. Birks and A. F. Lotter for valuable discussions and statistical advice. Many thanks to I. Renberg and I. Larocque for commenting on earlier drafts of the paper, and to A. Jonsson and J. Karlsson for providing pH and DOC data from their 16-lake set. K. Ericsson, E. Grahn, A. Jonsson, J. Karlsson, L. Janeck, P. Rosén, K. Sjödin and T. Westin assisted with lake sampling and lab support. We thank P. R. Leavitt and M. Graham for measuring LOI of our training-set lake sediments and K. Aune for creating Figure 1. The manuscript was written while RIH was employed at the University of Waterloo.

References

Alexandersson, H., C. Karlström & S. Larsson-McCann, 1991. Temperature and precipitation in Sweden 1961–1990. Reference normals. SMHI-Report No. 81, Norrköping, 88 pp.

Anderson, N. J., 2000. Diatoms, temperature and climatic change. Eur. J. Phy. 35: 307–314.

Battarbee, R. W., 1986. Diatom analysis. In Berglund, B. E. (ed.), Handbook of Holocene Paleoecology and Paleohydrology. John Wiley & Sons Ltd., London, 527–570.

Battarbee, R. W., 2000. Palaeolimnological approaches to climate change, with special regard to the biological record. Quat. Sci. Rev. 19: 107–124.

Bigler, C., R. I. Hall & I. Renberg, 2000. A diatom-training-set for paleoclimatic inferences from lakes in northern Sweden. Verhand. Int. Ver. Limnol. 27: 1174–1182.

Birks, H. J. B., 1995. Quantitative paleoenvironmental reconstructions. In Maddy, D. & J. S. Brew (eds), Statistical Modelling of Quaternary Science Data. Cambridge Quaternary Research Association, Cambridge XII, 161–254.

Birks, H. J. B., 1998. Numerical tools in paleolimnology – progress, potentialities, and problems. J. Paleolim. 20: 307–332.

Birks, H. J. B., J. M. Line, S. Juggins, A. C. Stevenson & C. J. F. ter Braak, 1990. Diatoms and pH reconstruction. Phil. Trans. R. Soc. Lond. B327: 263–278.

Brooks, S. J. & H. J. B. Birks, 2000. Chironomid-inferred late-glacial and early-Holocene mean July air temperatures for Kråkenes Lake, western Norway. J. Paleolim. 23: 77–89.

Camburn, K. E. & J. C. Kingston, 1986. The genus *Melosira* from soft-water lakes with special reference to northern Michigan, Wisconsin and Minnesota. In Smol, J. P., R. W. Battarbee, R. B. Davis & J. Meriläinen (eds), Diatoms and Lake Acidity. Dr. W. Junk Publishers, Dordrecht, 17–34.

Davis, R. B., D. S. Anderson & F. Berge, 1985. Paleolimnological evidence that lake acidification is accompanied by loss of organic matter. Nature 316: 436–438.

Dean, W. E., 1974. Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. J. Sed. Petrol. 44: 242–248.

De Nicola, D. M., 1996. Periphyton responses to temperature at different ecological levels. In Stevenson, R. J., M. L. Bothwell & R. L. Lowe (eds), Algal Ecology: Freshwater Benthic Ecosystems. Academic Press, San Diego, 149–181.

Felzer, B., S. T. Thompson, D. Pollard & J. C. Bergengren, 2000. GCM-simulated hydrology in the Arctic during the past 21,000 years. J. Paleolim. 24: 15–28.

Hall, R. I. & J. P. Smol, 1996. Paleolimnological assessment of long-term water-quality changes in south-central Ontario lakes affected by cottage development and acidification. Can. J. Fish. Aquat. Sci. 53: 1–17.

Heiri, O., A. F. Lotter & G. Lemke, 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. J. Paleolim. 25: 101–110.

Hill, M. O. & H. G. Gauch, 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio 42: 47–58.

Holmgren, B. & M. Tjus, 1996. Summer air temperatures and tree line dynamics at Abisko. Ecol. Bull. 45: 159–169.

Huisman, J., H. Olff & L. F. M. Fresco, 1993. A hierarchical set of models for species response analysis. J. Veg. Sci. 4: 37–46.

Hustedt, F., 1942. Diatomeen aus der Umgebung von Abisko in Schwedisch-Lappland. Arch. Hydrobiol. 39: 82–174.

Juggins, S., 1997. CALIBRATE version 0.70. A C++ program for analysing and visualising species environment relationships and for predicting environmental values from species assemblages. User guide Version 1.0. Department of Geography, Newcastle.

Korhola, A., J. Weckström, L. Holmström & P. Erästö, 2000. A quantitative Holocene climate record from diatoms in northern Fennoscandia. Quat. Res. 54: 284–294.

Korhola, A., J. Weckström & M. Nyman, 1999. Predicting the long-term acidification trends in small subarctic lakes using diatoms. J. Appl. Ecol. 36: 1021–1034.

Krammer, K. & H. Lange-Bertalot, 1986-1991. Bacillariophyceae, vol. 2(1–4). In Ettle, H., J. Gerloff, H. Heyning & D. Mollenhauer (eds), Süsswasserflora von Mitteleuropa, Gustav Fischer Verlag, Stuttgart/Jena.

Laaksonen, K., 1976. The dependence of mean air temperatures upon latitude and altitude in Fennoscandia (1921–1950). Ann. Acad. Sci. Fenn 119 (Series A III): 5–19.

Larocque, I., R. I. Hall & E. Grahn, in press. Chironomids as indicator of climate change: A 100-lake training set from a subarctic region of northern Sweden (Lapland). J. Paleolim.

Livingstone, D. M. & A. F. Lotter, 1998. The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palaeolimnological implications. J. Paleolim. 19: 181–198.

Lotter, A. F., H. J. B. Birks, W. Hofmann & A. Marchetto, 1997. Modern diatom, cladocera, 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.

Lotter, A. F., R. Pienitz & R. Schmidt, 1999. Diatoms as Indicators of Environmental Change near Arctic and Alpine Treeline. In Stoermer, E. F. & J. P. Smol (eds), The Diatoms: Application to the Environmental and Earth Sciences. Cambridge University Press, Cambridge, 205–226.

MacDonald, G. M., B. Felzer, B. P. Finney & S. L. Forman, 2000. Holocene lake sediment records of Arctic hydrology. J. Paleolim. 24: 1–14.

Olander, H., H. J. B. Birks, A. Korhola & T. Blom, 1999. An expanded calibration model for inferring lakewater and air temperatures from fossil chironomid assemblages in northern Fennoscandia. The Holocene 9: 279–294.

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.

Renberg, I., 1990. A procedure for preparing large sets of diatom slides from sediment cores. J. Paleolim. 4: 87–90.

Renberg, I., 1991. The HON-Kajak sediment corer. J. Paleolim. 6: 167–170.

Renberg, I., T. Korsman & H. J. B. Birks, 1993. Prehistoric increases in the pH of acid-sensitive Swedish lakes caused by land-use changes. Nature 362: 824–826.

Rosén, P., R. I. Hall, T. Korsman & I. Renberg, 2000a. Diatom transfer-functions for quantifying past air temperature, pH and total organic carbon concentration from lakes in northern Sweden. J. Paleolim. 24: 109–123.

Rosén, P., E. Dåbakk, I. Renberg, M. Nilsson & R. Hall, 2000b. Near-infrared spectrometry (NIRS): a new tool for inferring past climatic changes from lake sediments. The Holocene 10: 161–166.

Smol, J. P. & B. F. Cumming, 2000. Tracking long-term changes in climate using algal indicators in lake sediments. J. Phycol. 36: 986–1011.

Smol, J. P., I. R. Walker & P. R. Leavitt, 1991. Paleolimnology and hindcasting climatic trends. Verh. Int. Ver. Limnol. 24: 1240–1246.

Stevenson, A. C., H. J. B. Birks, R. J. Flower & R. W. Battarbee, 1989. Diatom-based pH reconstruction of Lake Acidification Using Canonical Correspondence Analysis. Ambio 18: 228–233.

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

ter Braak, C. J. F., S. Juggins, H. J. B. Birks & H. van der Voet, 1993. Weighted averaging partial least squares regression (WA-PLS): definition and comparison with other methods for species-environment calibrations. In Patil, G. P. & C. R. Rao (eds), Multivariate Environmental Statistics. Elsevier Science Publishers, Amsterdam, 525–560.

ter Braak, C. J. F. & I. C. Prentice, 1988. A theory of gradient analysis. Adv. Ecol. Res. 18: 271–317.

ter Braak, C. J. F. & P. Smilauer, 1998. CANOCO Reference Manual and User´s Guide to Canoco for windows: Software for Canonical Community Ordination. Ithaca, NY, Microcomputer Power, 351 pp.

Weckström, J., A. Korhola & T. Blom, 1997. Diatoms as quantitative indicators of pH and water temperature in subarctic Fennoscandian lakes. Hydrobiologia 347: 171–184.

Weller, G. & M. Lange, 1999. Impacts of Global Climate Change in the Arctic Regions. Tromsø, IASC: 1–59.

Wunsam, S., R. Schmidt & R. Klee, 1995. *Cyclotella*-taxa (Bacillariophyceae) in lakes of the Alpine region and their relationship to environmental variables. Aquat. Sci. 57: 360–386.

Vyverman, W. & K. Sabbe, 1995. Diatom-temperature transfer functions based on the altitudinal zonation of diatom assemblages in Papua New Guinea: a possible tool in the reconstruction of regional palaeoclimatic changes. J. Paleolim. 13: 65–77.