ENVIRONMENTAL CHANGE RESEARCH CENTRE

University College London

RESEARCH REPORT

No. 33

Palaeolimnology and Lakes with respect to pollution and climate change

Annual Report, Year 2 (1995 - 1996)
INTAS Contract: 1010-CT93-0021

Solovieva, N. & Patrick, S.T.

November 1996

Environmental Change Research Centre
University College London
26 Bedford Way
London
WC1H 0AP

OVERVIEW

The objectives of this INTAS programme between the Environmental Change Research Centre (ECRC), University College London; Institute of Global Climate and Ecology (IGCE) Moscow; Moscow State University (MSU) Department of Hydrobiology; and the Kola Science Centre (KSC) Apatity, are to introduce recently developed palaeolimnological methods to laboratories in the Former Soviet Union and to apply these techniques collaboratively to problems of environmental change and pollution. The focus of this work is centred on atmospheric pollution and potential climate change in the Kola Peninsula and the Lake Baikal region.

In the contract year 1995-1996, young scientists from ECRC have visited IGCE and MSU to discuss ideas and the KSC to undertake fieldwork and laboratory analyses of sample material. Senior scientists from IGCE, MSU and KSC have visited ECRC to review the collaborative programme and young scientists from the three participating Russian laboratories have attended courses in numerical analysis and diatom taxonomy at the ECRC.

Inevitably the parlous financial situation in Russian academic institutions has affected the progress of this project. The science programme in the Baikal region has been difficult to maintain and the emphasis in 1995-1996 has been placed on the Kola Peninsula. The logistics of fieldwork and laboratory back-up have been easier to arrange with the involvement of KSU, but even here economic difficulties have precluded certain analyses being undertaken. Additionally, as salaries for senior staff and financial support for young scientists have virtually ceased to exist it is inevitable that money for some of the equipment and consumable purchases designated in the original proposal, has been diverted to help maintain the position of key researchers involved with this project.

As part of the palaeoecological reconstruction of recent environmental (acidification) and post Holocene climate history of the Kola Peninsula, surface sediments and water samples have been collected from 27 lakes located throughout the Kola Peninsula along the vegetation gradient. Two lakes were subsequently excluded from the training set due to disturbance in the catchments. The final training set thus contains 25 sites. The background to this study and the first results are presented in the scientific report below.

PALAEOLIMNOLOGICAL ASSESSMENT OF CLIMATE CHANGE AND ACIDIFICATION OF LAKES IN THE KOLA PENINSULA

1 INTRODUCTION

The role of palaeoenvironmental data in understanding the natural variability of the global climate and possible future climate change has been widely recognised because palaeodata provide a 'window on the past' and allow an estimation of how the Earth climate system functioned under different conditions. Palaeoenvironmental records facilitate opportunities for hypothesis-testing and validation of General Circulation Models developed for assessment and prediction of global climate variations (PAGES, 1995).

Lake sediments gradually accumulate physical, chemical and biological material derived from the lake, its catchment and the atmosphere and hence the environmental history of the lake and its catchment can be revealed by sedimentary analysis (Battarbee, 1992).

The environmental and climate history of the Kola Peninsula has attracted especial scientific interest for the following reasons:

- The Arctic environment has been found to be particularly vulnerable to even slight changes in environmental conditions and thus environmental change is more pronounced there (Cwynar & Spear, 1991) and may be more easily traced. High latitude regions are also relatively pristine and still remote from human disturbance therefore providing an opportunity to obtain data free from anthropogenic impact.

- Possible future global warming is expected to particularly effect Arctic coastal regions causing changes in ecotonal forest/tundra boundaries (i.e. tree-line). Redistribution of boreal forest may in turn affect the global climate. It is therefore important to establish how northern ecosystems have responded to rapid changes in climate, which occurred during the Holocene (Pienitz & Smol, 1993) because such climate events provide an analogue for the impact of future global change on Arctic ecosystems (MacDonald *et al.*, 1993). The timber-line also indicates the mean summer location of the Arctic Front and hence determining its position in the past will be important for climatological predictive models.

The impact of the Holocene climate change on the boreal forest/tundra ecotone has been extensively studied in Canada (MacDonald *et al.*, 1993; Payette & Lavoie, 1994; Pienitz *et al.*, 1995); and similar projects are being conducted in Scandinavia under the Nordchill project (Seppa, 1996). The PALE project (Palaeoclimate from Arctic Lakes and Estuaries), which is part of the PAGES (Past Global Changes) International Geosphere-Biosphere Programme, will include the palaeoclimatic studies of Beringia (Eastern Alaska-Siberia).

Therefore, the study of Holocene environmental change in Russian arctic ecosystems will give a more accurate account of the Holocene climate events in the Northern Hemisphere. It will also contribute a general understanding of circumpolar climate change in the Northern Hemisphere and therefore will help to predict and to assess future climate change.

The use of pollen and diatom analysis in palaeoenvironmental reconstructions

In general, several palaeoecological methods are usually applied to derive climate reconstructions. Past climate conditions can be assessed by reconstructing the vegetation of the area using pollen analysis. However, the low pollen production of the Arctic regions (Birks, 1991), the high influence of long distance extra-regional transport of pine pollen (Seppa, 1996) and peculiarities of pollen dispersal and pollen representation in the sediment, might bias the results of pollen based climate reconstructions and smaller scale climate events may be missed out. According to Wright (1984), and Davis & Botkin (1985) pollen analysis alone may not provide enough information for reconstructing the dynamics of plant communities in northern latitudes and other palaeoecological indicators will substantially enhance analysis (e.g. Smol *et al.*, 1991).

Lake diatom assemblages can possibly reflect climate dynamics both in direct and indirect ways. Results of several studies (e.g. Smol *et al.*, 1991; Pienitz & Smol, 1993; Pienitz *et al.*, 1995) imply that a direct relationship between diatom species composition and temperature or latitudinal gradient exists. However, this evidence can be also an indicator of diatom response to the change of microhabitat and water chemistry caused by temperature shift.

Transition of species composition, species diversity and abundance of the diatom community reflect various environmental changes within the lake/catchment ecosystem which are driven by climatic transition. Climate change can lead to a change of erosion rates and an influx of mineral and organic matter from the drainage basin (Mackereth, 1966) which in turn will affect lake chemical and physical characteristics (e.g. transparency, degree of light penetration, pH, dissolved organic carbon, calcium and sodium concentrations). This subsequently influences the composition of diatom assemblages (Pienitz *et al.*, 1995) and finally these changes are recorded in sediments by frustules of fossil diatoms.

Thickness of ice and snow cover of a lake also depends on climate as well as the length of ice-free period and growing season. These environmental variables control productivity, growth, photosynthetic activity and therefore species composition of diatom assemblages (e.g. Tilman & Kiesling, 1984; Rhee & Gotman 1981).

Given the high northern latitude location of the study area, remote from human disturbance, featuring bedrock with low buffering capacity (mainly granite and gneiss), thin unproductive soils, dispersed vegetation and low productivity lakes, diatom analysis together with geochemical analyses of lake sediments, were chosen as the most appropriate palaeoecological techniques for climate reconstruction in addition to pollen analysis.

Acidification history and climate

The indirect link between climate change and change in lake acidity has been revealed only recently (Schindler *et al.*; 1996, Psenner & Schmidt, 1992). As discussed above, climate transition also causes changes in the catchment erosion and leaching rates which influences water chemistry and turbidity. The influx concentration of allogenic organic matter (dissolved organic carbon - DOC) is controlled by the type of vegetation cover and soils of the catchment, which are climate related parameters (Pienitz & Smol, 1993). For example, it was found that the level of all nutrients and major ions are higher in the lakes of the boreal forest

zone of Canada than in the tundra zone. Many studies reveal the strong correlation between DOC and diatom distribution (e.g. Stevenson *et al.*, 1989; Kingston & Birks, 1990) which provides another indirect link between diatoms and climate.

On the other hand, the decrease in pH is usually associated with decline in DOC (e.g. Davis *et al.*, 1985; Stevenson *et al.*, 1989) and pH is an important factor controlling diatom distribution (Battarbee, 1986; Dixit *et al.*, 1991). However, in North American boreal lakes acidification was accompanied by decline in DOC and was related to climate warming during recent decades (Schindler *et al.*, 1996). The dryer conditions as a consequence of climate warming in this region initially caused a reduced influx of DOC into the lakes and then lake productivity also declined.

The other interesting finding of the work of Schindler *et al.* (1996) is that the decline in DOC, especially in its coloured allochthonous part which is a function of catchment/lake ratio, leads to deeper ultraviolet penetration into the water column. This can be potentially harmful for aquatic biota and account for major ecosystem changes (e.g. changes in floristic composition) which can be later recorded in sediments. Boreal clear water dilute lakes are rather susceptible to this ultraviolet impact. Since the primary cause of climate change is often considered to be the fluctuations in solar radiation, lake sediments might well reflect these shifts in solar activity.

Alternatively, the studies of acidification of European Alpine lakes (Psenner & Schmidt, 1992) revealed the opposite: i.e. that warmer (and dryer) periods were associated in this region with higher pH in comparison with colder (and dryer years). This was attributed to increased in-lake and catchment alkalinity production and more intensive weathering of bedrock during the warmer periods.

Accordingly, reconstruction of lake acidification history in the Kola region using diatom analysis, will help reveal the climate changes occurred in the past.

2. MODERN DIATOM DATASET AND DEVELOPMENT OF DIATOM-pH TRANSFER FUNCTION

This study focuses on an exploration of the relationship between modern diatom sediment composition and chemical, physical and vegetational characteristics of lakes and their catchments. The main objectives of this part of the project are to identify the environmental variables influencing diatom distribution in the study lakes of the Kola Peninsula, to assess the relative effect of climate-related variables on diatom composition and to develop an appropriate transfer function for later palaeoecological reconstructions. A series of explorative multivariate statistical methods are applied to estimate the significance of different environmental variables.

Sites

The general characteristics of the geography, geology, climate and vegetation of the study area were given in the First Progress Report.

The geographic distribution of sampling sites is shown in Figure 1. The summary geographical characteristics of lakes and their catchments are presented in Appendix 1. Most of the lakes studied have no names. They were therefore numbered in the order of sampling and they are prefixed 'KOLA' on the ECRC database.

Materials and methods

Field and laboratory methods followed those detailed in (Pienitz, *et al.*, 1995) and the protocols recommended under the PAGES initiative, PALE and those adopted by the AL:PE programme (Cameron *et al.*, 1993).

Lake altitudes, lakes areas and quantitative catchment characteristics were derived from topographic maps. Catchment vegetation and geology were described in the field. The descriptions were also based on the *Atlas of Murmansk region* (1971).

The fieldwork was organized by the Kola Science Centre and was carried out from 14 August to 3 September 1995.

Samples of surface sediments were collected using the standard Glew corer (Glew, 1991). The sediments were collected from the deepest point of the lake determined with a portable echosounder. All cores were extruded in the field using standard extrusion equipment. The top 1 cm of sediment, which represents an integrated sample (in space and time) of the diatoms over the previous few years were used in this study.

Surface-water samples were collected from the centre of the same lakes from a depth of 0.5 m.

Water samples were analyzed for 18 chemical parameters: pH, conductivity (COND), water colour (COL), ammonium (NH_4), phosphate (PO_4), calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), sulphate (SO_4), chloride (Cl), silica (Si), total nitrogen (N_{tot}), total phosphorus (P_{tot}), alkalinity (ALK), chemical oxygen demand (COD), total organic carbon (TOC) and iron (Fe). The analysis was carried out in the Aquatic Ecosystem Laboratory, Kola Science Centre using standard techniques (Rond *et al.*, 1975). Chemical data quality are ensured through participation of the Laboratory in the Intercalibration Programme administered by the UN-ECE International Cooperative Programme on Acidification of Lakes and Rivers.

Preparation of diatom slides followed techniques outlined in Battarbee (1986). Slides were mounted using Naphrax. A minimum of 500 valves were identified and counted per slide, using a Zeiss light microscope with magnification of x1000 using oil immersion. Diatoms were identified to the lowest possible taxonomic level following Krammer and Lange-Bertalot guides (1986-1991).

Environmental dataset.

Chemical variables (mean data) used for subsequent analysis are shown in Appendix 2. Diatom taxa that were found in at least three lakes with abundance > 1% were included into the dataset. By this means, the total dataset comprises 24 environmental variables and 99

diatom taxa.

Table 1 compares the results of the summary statistics (maximum, minimum and mean) of water chemistry for the total training set and for the different catchment vegetational zones i.e. tundra, forest-tundra and boreal forest, and two lakes separately classified in alpine zones. Table 2 presents the summary statistics of the environmental characteristics of the lakes and their catchments.

On the whole, water chemistry characteristics of the total set of lakes are very diverse. The diversity of chemical parameters of the study sites is also indicated in Figure 2 which shows frequency histograms of a number of chemical variables. The correlations between several environmental variables are shown in Figure 3.

pH and alkalinity

The pH data have skewed distribution with the majority of values falling in the range 6.3 to 6.7 with a mean for the whole training set of 6.35. pH values for the full dataset range from 5.0 to 7.44. Only four sites (KOLA3, KOLA17, KOLA19 and KOLA22) have pH values lower than 6.0 and two sites (KOLA1 and KOLA2) have pH values greater than 7.0.

Alkalinity values are generally low with a mean for the dataset of 72.77 $\mu eq\ l^{-1}$. All sites fall in the range from -9 to 283 $\mu eq\ l^{-1}$. As expected, the four lowest pH sites have low alkalinity and are the only sites with negative alkalinity values. Five sites have a relatively high alkalinity of over 100 $\mu eq\ l^{-1}$ (KOLA1, KOLA2, KOLA13, KOLA15, KOLA18 and KOLA20). Site KOLA2 (alpine forest/tundra) has the highest alkalinity value within the training set because it situated in the Khibiny Mountains with ultra-alkaline bedrock.

Both pH and alkalinity show no geographical patterns in the distribution of their values among the lakes within the training set. Generally, tundra lakes have a smaller range of pH and alkalinity compared to lakes of other vegetational zones (Figure 4).

Conductivity

All lakes are dilute. Conductivity values range from 8.00 to 88.00 $\mu S\ cm^{-1}$ with most sites falling in the range 20-42 $\mu S\ cm^{-1}$ and with a mean value of 32.74 $\mu S\ cm^{-1}$. Site KOLA6 (forest/tundra zone) has the maximum conductivity value within the training set of 88 $\mu S\ cm^{-1}$, which may be due to its proximity to the sea coast. Other sites located in the vicinity of the sea (e.g. KOLA23, KOLA24) also have rather high conductivity values (see Appendix 2). Sites KOLA2 and KOLA1 also have high conductivity values compared to other lakes because of their ion-rich geology.

Sites KOLA25 (alpine tundra) and KOLA17 (boreal forest) are characterized by the lowest conductivity values within the training set (9 $\mu S\ cm^{-1}$ and 8 $\mu S\ cm^{-1}$ respectively).

Major ions

There is high variation in the concentrations of all ions. Sites KOLA6 and KOLA2 are characterized by the highest values for all major ions for the reasons discussed above. The

lowest water concentrations of all major ions were expectedly found in KOLA17 and KOLA25.

All lakes have low potassium concentrations ranging from 0.04 (KOLA25) to 1.17 mg l^{-1} (KOLA2) and the mean for the whole training set of 0.394 mg l^{-1}. The potassium concentrations of the majority of sites (20 out of 25) vary within the narrow range from 0.1 to 0.5 mg l^{-1}. Calcium concentrations for all lakes are also comparatively low varying from 4.18 mg l^{-1} (KOLA1) to 0.3 mg l^{-1} (KOLA17) with a mean of 1.4 mg l^{-1}. Sixteen lakes have calcium concentrations between 0.3 and 1.5 mg l^{-1}.

Magnesium concentrations form a normal distribution (Figure 2) i.e. the greater number of sites fall into a middle-range between 0.55 and 0.9 mg l^{-1} (the mean is 0.741 mg l^{-1}). The site KOLA13 (boreal forest) has the maximum magnesium concentration of 1.4 mg l^{-1} and KOLA17 has the lowest magnesium value of 0.14 mg l^{-1}.

Calcium and potassium concentrations are closely correlated and show a similar type of distribution within the training set. These ions show similarities in their distribution to pH and alkalinity.

No clear correlation between vegetation zones and distribution of potassium, calcium and magnesium was observed, although tundra lakes have lower values of calcium and potassium (see Table 1) compared to lakes with afforested catchments. The range of calcium and potassium is also smaller in tundra sites. This can be influenced by the generally low level of alkaline ions in tundra soils. Magnesium concentrations show no difference between tundra and forest zones.

Sodium concentrations vary greatly within the training set and approximate a normal distribution (Figure 2). The mean sodium concentration for the whole training set is 3.53 mg l^{-1} and 11 lakes fall into the middle range with concentrations from 3 to 5 mg l^{-1}. Maximum sodium concentration was found in the forest-tundra site KOLA6 (11.3 mg l^{-1}) and the lowest value of 0.44 mg l^{-1} was at KOLA17.

Chloride concentrations show the greatest range - from 0.62 mg l^{-1} in the alpine tundra KOLA25 to 21.9 mg l^{-1} in KOLA6. The distribution of chloride concentrations is negatively skewed, i.e. 10 lakes have chloride values below 2.0 mg l^{-1} and the chloride concentrations of the other 15 lakes are almost evenly distributed within the range from 2 to 10 mg l^{-1}.

Sulphate distribution (Figure 2) varies from 1.06 mg l^{-1} (KOLA17) to 5.0 mg l^{-1} (KOLA1 and KOLA19, forest sites) with a mean of 2.48 mg l^{-1}. Inland sites KOLA1 and KOLA19 may be slightly influenced by air pollution and therefore exhibit higher sulphate values

Sodium and chloride concentrations are closely intercorrelated and much influenced by sea salts. This is shown in Figure 3 by the positive correlation with latitude and longitude, reflecting the distance from the sea and concentrations of these ions. The highest concentrations of sodium and chloride are found in the tundra and forest/tundra sites which are closer to the sea than the boreal forest sites.

Distribution of sulphate concentrations in the lakes seems not to be influenced by seasalts and shows no clear geographical patterns. However, there is a positive correlation between sulphate and sodium/chloride (Figure 3). The mean for the forest zone is the highest, and the highest values of sulphate are also found within the boreal forest. This may well reflect the influence of air pollution from the large smelters of Monchegorsk.

Nutrients

All lakes have a low level of nutrients and are of oligotrophic or dystrophic status which is typical for this geographical region.

The total nitrogen concentrations vary from 470 μg l^{-1} (KOLA3, forest/tundra) to 94 μg l^{-1} (KOLA8, tundra) with a mean value of 226 μg l^{-1}. The distribution of nitrogen concentrations shows two distinctive maxima. Nitrogen values for six lakes fall into the range from 100 to 150 μg l^{-1} and nine lakes have nitrogen concentrations from 200 to 250 μg l^{-1} (Figure 3).

Phosphorus values form a normal distribution with the majority of lakes having concentrations in a middle range from 4 to 8 μg l^{-1} (the mean for the total set is 7.44 μg l^{-1}). Phosphorus concentrations vary from 2 μg l^{-1} (KOLA24, tundra) to 17 μg l^{-1} (KOLA3, forest/tundra zone).

Nitrogen concentrations are almost two times higher in lakes with forest catchments, with the greatest values in the forest/tundra zone. Phosphorus concentrations are higher in the forest/tundra zone and they are similar in the tundra and forest zones. On the whole, forest/tundra lakes are the most diverse in terms of nutrient concentrations. Nitrogen and phosphorus water content are positively intercorrelated.

Organic matter

Three parameters indicating the level of organic matter in the lake water were measured: total organic carbon (TOC), water colour (COL) and chemical oxygen demand (COD). All three determinands are closely interrelated and show little difference in their distribution. Therefore only the range and distribution of TOC will be discussed.

The site KOLA25 (alpine tundra) has the lowest values for the three parameters and the site KOLA20 exhibits the maximum for dissolved organic matter.

Distribution of dissolved organic matter in the lakes has two major peaks. One maximum is formed by lakes with TOC concentrations between 3 and 6 mg l^{-1}. These nine lakes with very clear water and rocky and sandy catchments are located predominantly within the tundra zone, but there are also some lakes represented from the forest zone. Another maximum (TOC varies from 9 to 12 mg l^{-1}) is comprised by the tea-coloured lakes within peatland areas prevailing in the forest/tundra zone. Generally, forest/tundra lakes exhibit the greater diversity in organic matter concentrations compared to tundra and forest lakes (Table 1).

Morphometric characteristics

Measured limnological characteristics comprise: maximum water depth and lake area (Appendix 1, Table 2). Maximum depth ranges from 1.5 m to 19.2 m with a mean for the

total training set of 6.5 m. The majority of lakes fall in the range 3 to 7 m (Appendix 1, Table 2). The tundra lakes tend to be deeper (mean 7.8 m) than the forest and forest/tundra lakes (means 4.5 and 6 m respectively).

Lake area varies from 0.04 to 0.3 km^2. Most lakes are small in size (0.04 - 0.06 km^2). Two tundra lakes have the largest area within the training set and this influenced the mean for the tundra lake area (0.11 km^2 compared to 0.056 km^2 and 0.054 km^2 for the forest and forest/tundra zones respectively). On the whole, no geographical pattern in lake size was found.

Catchment characteristics

Catchment characteristics include percentage of forest and peat cover in the lake catchment which reflects the vegetation zones. Consequently, forest lakes have the highest percentage of forest and the tundra lakes have the lowest. Percentage of forest in the catchment varies from 0.1% to 92.9% throughout the whole training set.

Percentage of peatland area in the lake catchments varies from 0.1% to 46.8%. Forest lakes have the highest percentage of peat (26% on the average) and the tundra lakes have the lowest proportion of peatland area in their catchments (mean = 3.87%).

Geographical characteristics

Latitude, longitude and altitude comprise the geographical characteristics included into the set of environmental variables.

Latitude varies from 67°33' to 69°14', longitude - from 28°4' to 36°4' and altitude from 79 to 506 m. Geographical site location is related to vegetation zones and climate which was the principal reason to include these characteristics in the analysis. Latitude and especially longitude correlate well with the distance from the sea and reflect the influence of seasalts on diatoms.

Figure 1 **Site location map**

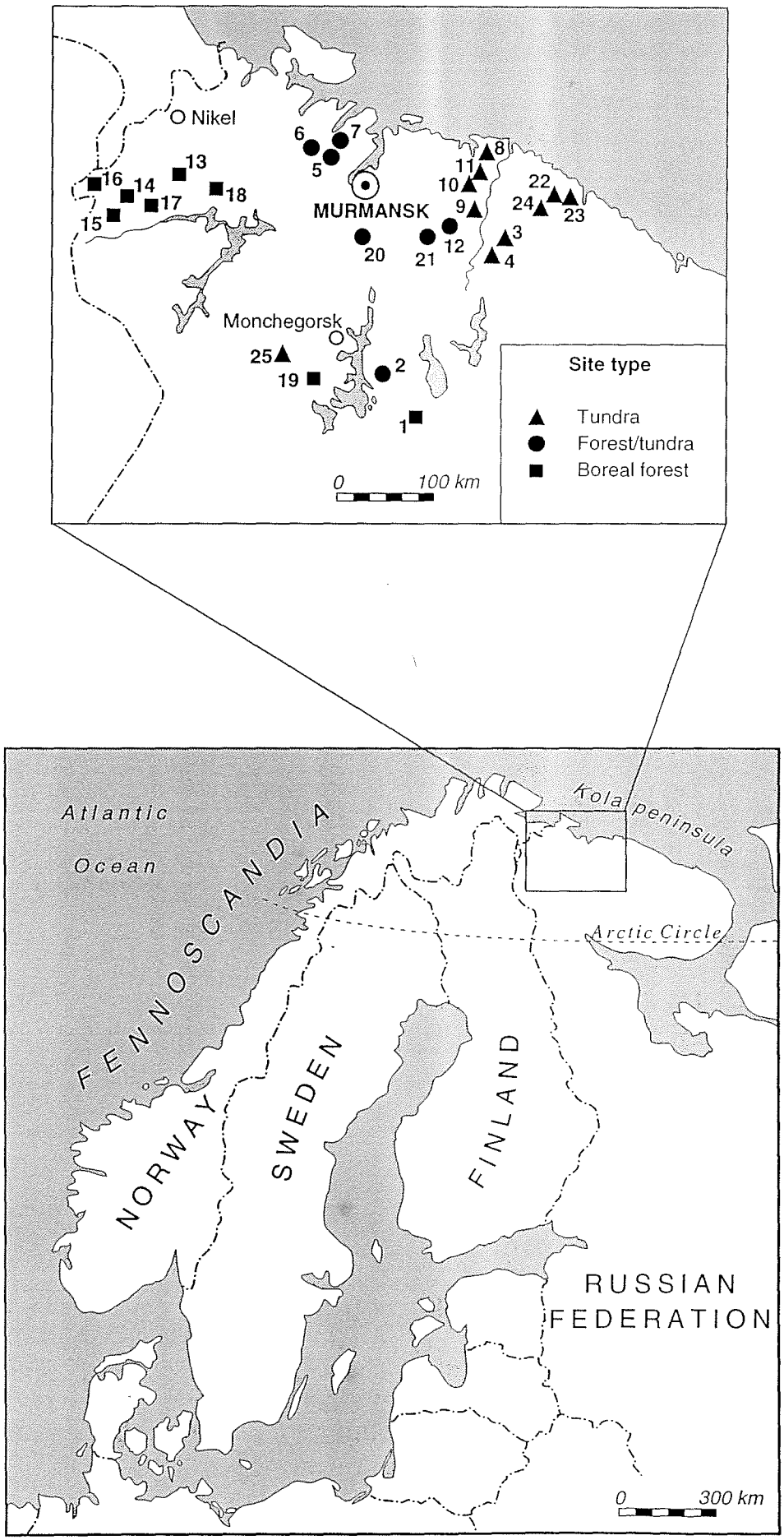
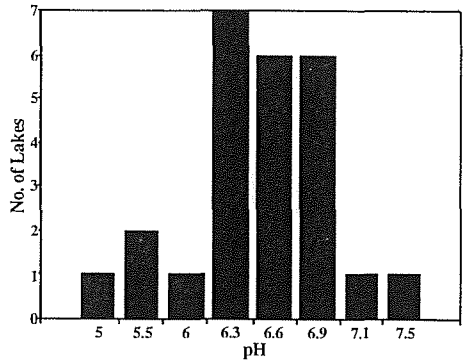
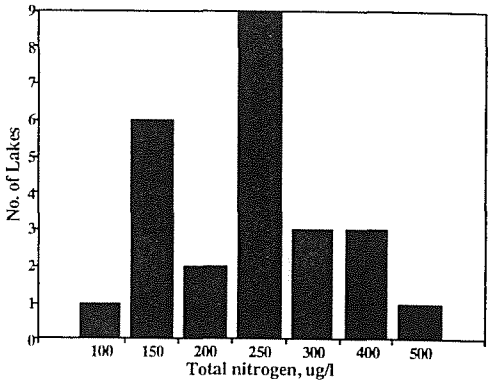
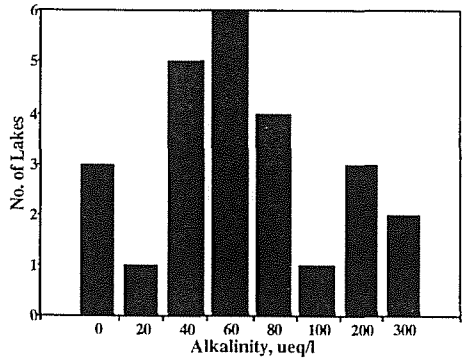
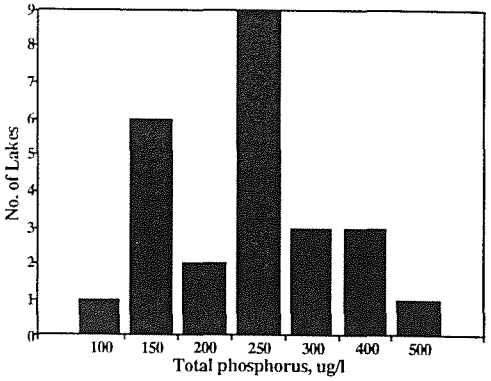


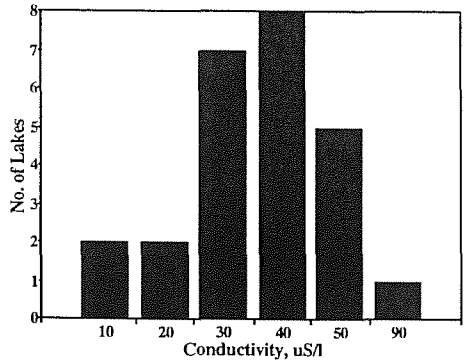
Figure 2 **Frequency histograms for selected chemical variables**











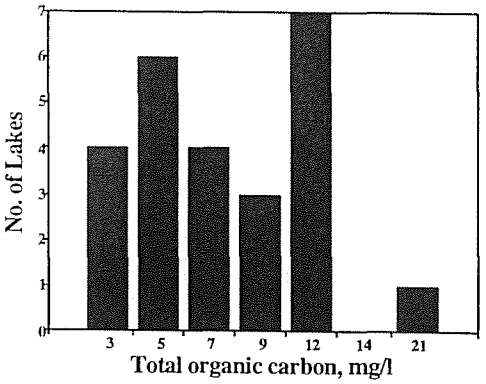
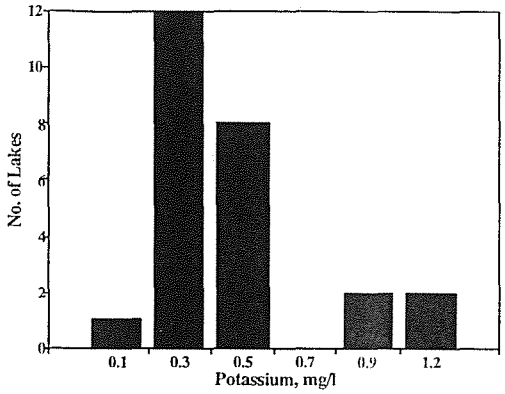
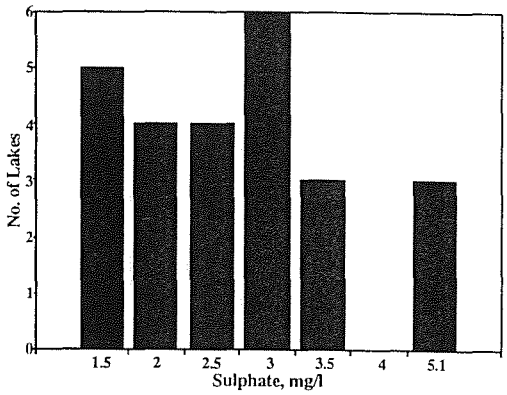
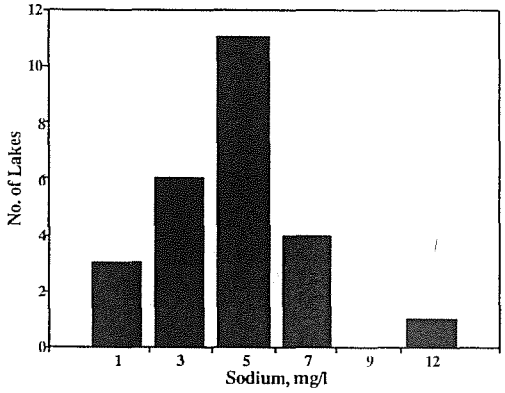
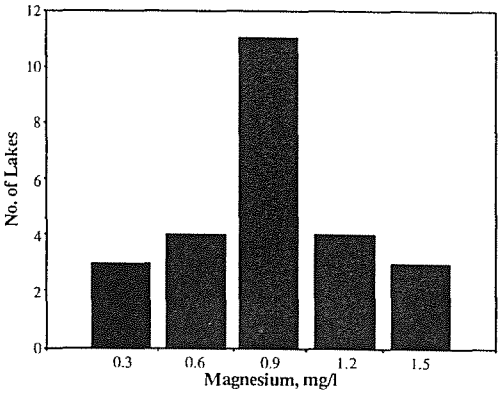


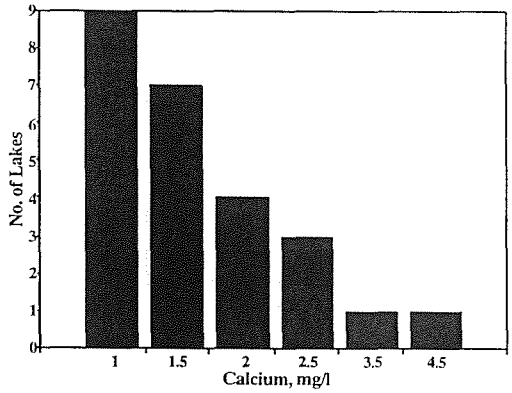
Figure 2 Cont.











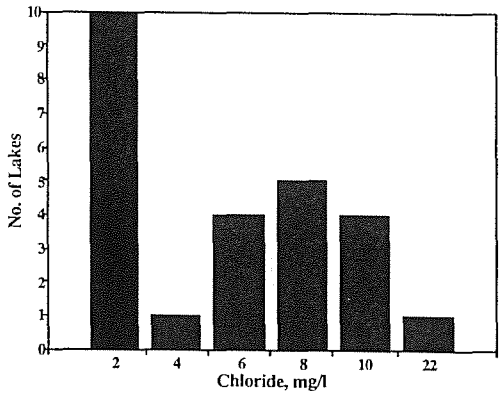


Figure 3 **Intercorrelations between selected environmental variables**

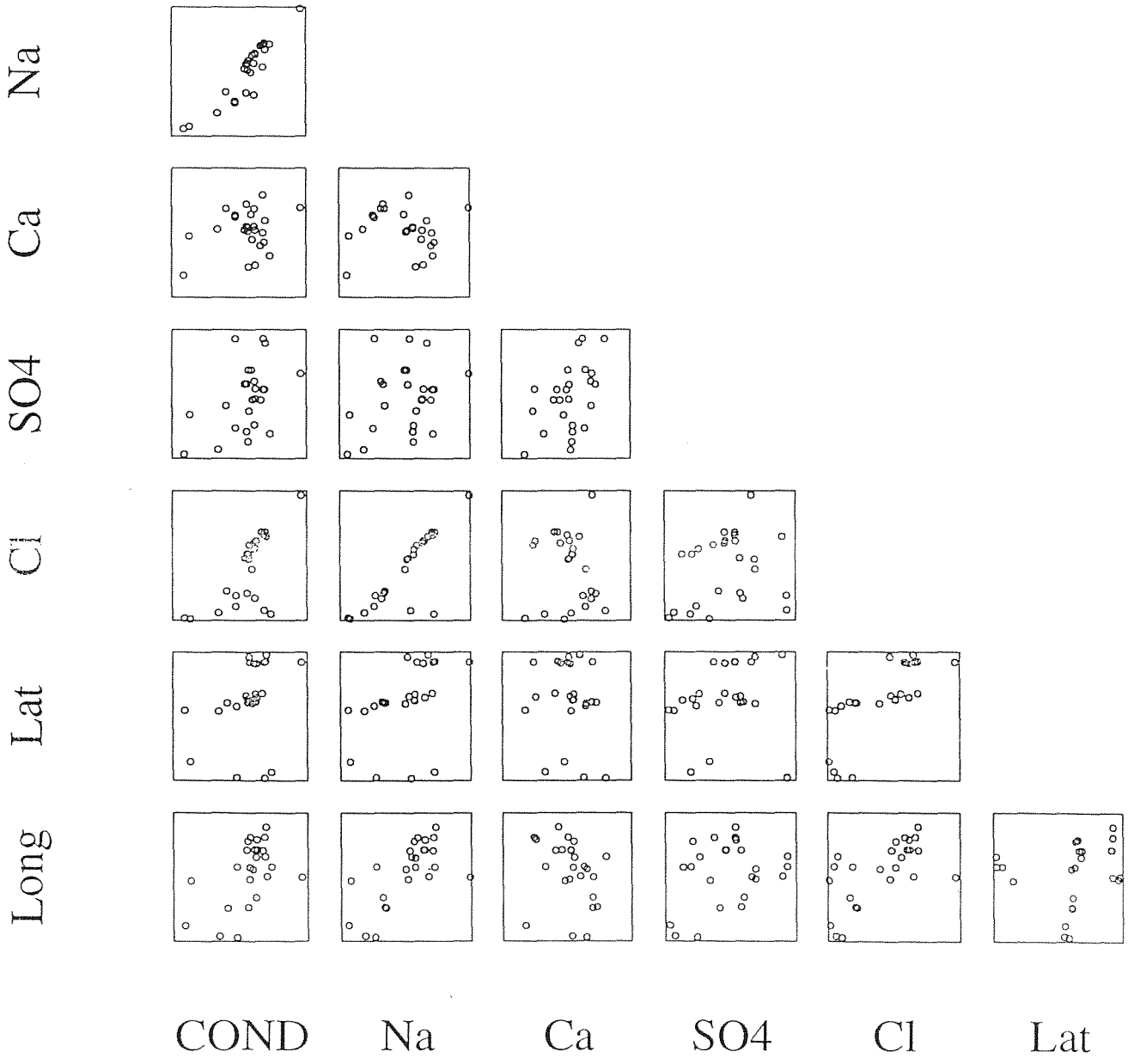


Figure 4 **Box and Whisker plots of pH values in different vegetation zones**

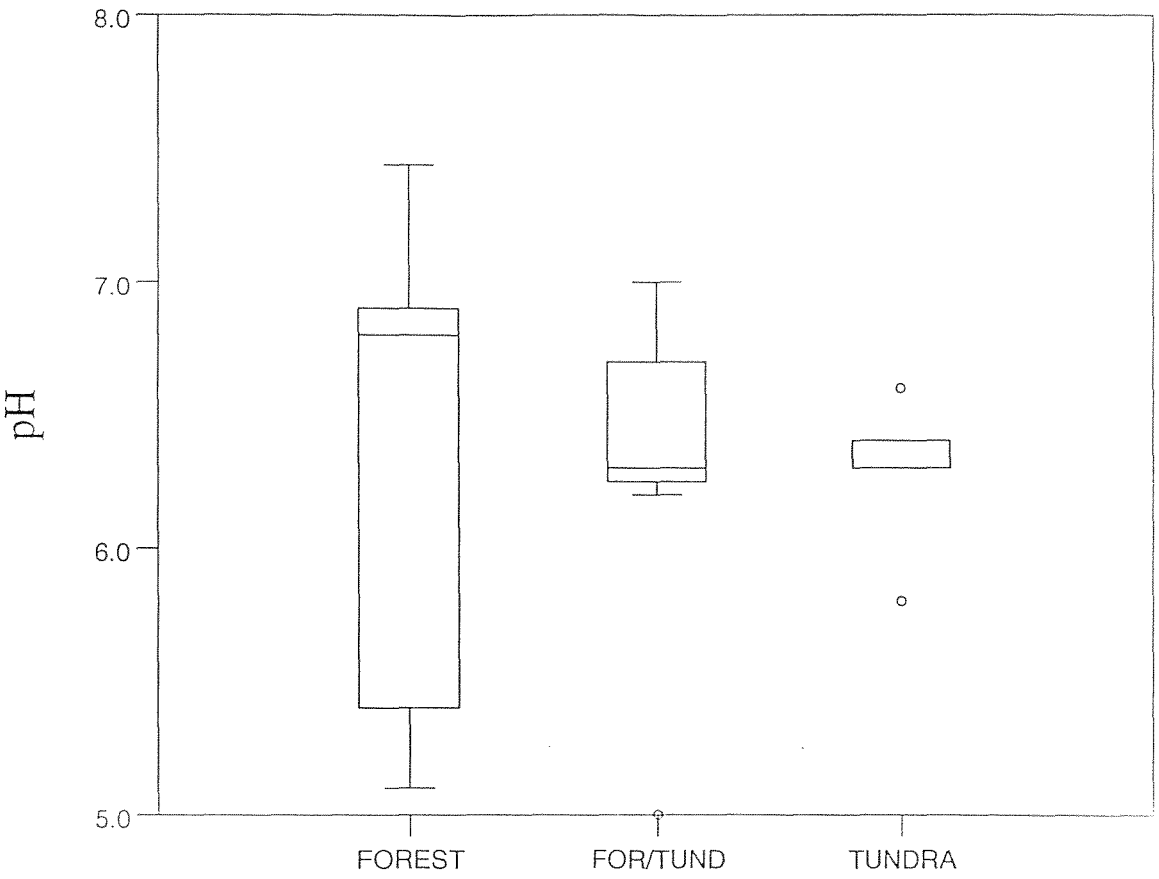


Table 1 Comparison of water chemistry data for 25 lakes located in different vegetation zones

Variable	Forest			Forest-tundra			Alpine forest-tundra	Tundra			Alpine tundra	All lakes combined		
	Min	Max	Mean	Min	Max	Mean		Min	Max	Mean		Min	Max	Mean
pH	5.1	7.44	6.41	5.0	7.0	6.32	7.0	5.8	6.6	6.32	6.4	5.0	7.44	6.35
COND ($\mu S\ cm^{-1}$)	8.0	41.0	24.1	29.0	88.0	40.88	47.0	9.0	42	32.06	9.0	8.0	88.0	32.7
ALK $\mu eq\ (l^{-1})$	-4.0	283.0	111.1	-9.0	210.0	69.0	210.0	24.3	59	36.53	24.3	-9.0	283.0	72.77
COL (°P)	3.0	120.0	57.87	5.0	240.0	98.55	5.0	16.0	46	33.0	4	3.0	240.0	63.4
NH_4 ($\mu g\ l^{-1}$)	3.0	21.0	9.0	1.0	24.0	10.33	15.0	3.0	28	10.07	7.0	1.0	28.0	9.84
N_{tot} ($\mu g\ l^{-1}$)	143.0	305.0	230.1	110.0	470.0	283.3	388.0	94.0	215	146.5	234.0	94.0	470.0	226.0
P_{tot} ($\mu g\ l^{-1}$)	4.0	12.0	6.75	3.0	17.0	8.55	10.0	2.0	10.0	6.14	12.0	2.0	17.0	7.44
COD ($mg\ l^{-1}$)	1.4	11.2	6.91	1.67	24.4	10.45	11.0	2.45	5.40	3.92	1.63	1.4	24.4	7.13
TOC ($mg\ l^{-1}$)	2.3	10.1	6.79	2.8	20.2	9.47	2.5	3.4	5.7	4.52	0.73	0.73	20.2	6.6
SiO_2 ($mg\ l^{-1}$)	0.04	2.37	0.87	0.16	2.91	1.26	2.9	0.53	1.12	0.66	0.7	0.04	2.91	0.94
K ($mg\ l^{-1}$)	0.13	0.99	0.46	0.23	0.83	0.53	1.17	0.23	0.31	0.27	0.04	0.04	1.17	0.39
Na ($mg\ l^{-1}$)	0.44	3.32	1.55	2.93	11.3	4.78	5.50	3.19	5.64	4.60	0.50	0.44	11.3	3.53
Ca ($mg\ l^{-1}$	0.3	4.18	2.11	0.39	2.48	1.27	0.50	0.37	1.22	0.85	0.90	0.3	4.18	1.4
Mg ($mg\ l^{-1}$)	0.14	1.40	0.68	0.63	1.36	0.89	1.00	0.56	0.89	0.69	0.14	0.14	1.4	0.74
SO_4 ($mg\ l^{-1}$)	1.06	5.0	2.66	1.26	4.71	2.51	1.40	1.90	2.71	2.33	1.80	1.06	5.0	2.48
Cl ($mg\ l^{-1}$)	0.65	1.94	1.31	3.71	21.9	13.5	0.80	4.96	9.43	7.74	0.62	0.62	21.9	5.21
Fe ($\mu g\ l^{-1}$)	1.8	190.0	64.97	19.0	320.0	124.3	2.0	20.0	49.0	34.62	49.0	1.8	320.0	72.39

Table 2 Comparison of environmental data for 25 lakes located in different vegetation zones

Variable	Forest			Forest-tundra			Alpine forest-tundra	Tundra			Alpine tundra	All lakes combined		
	Min	Max	Mean	Min	Max	Mean		Min	Max	Mean		Min	Max	Mean
Latitude	67.33	68.44	68.12	68.42	69.14	68.78	67.42	69.02	69.11	69.06	67.57	67.33	69.14	68.47
Longitude	28.41	34.05	30.65	32.37	35.37	33.87	33.28	34.52	36.04	35.28	32.27	28.41	36.40	32.99
Altitude (m)	104.0	179.0	128.0	101.0	240.0	170.5	460.0	79.0	260.0	155.87	506.0	79.0	506.0	174.4
Max Depth (m)	3.0	6.2	4.5	1.5	19.2	6.0	8.0	1.5	18.5	7.81	16.0	1.5	19.2	6.5
Area (km^2)	0.04	0.11	0.05	0.04	0.09	0.06	0.04	0.04	0.3	0.11	0.04	0.04	0.30	0.07
% forest	42.90	92.9	64.48	0.1	57.4	20.8	0.1	0.1	1.0	0.19	0.1	0.1	92.90	28.33
% peat	8.20	46.8	26.82	0.1	14.6	4.66	0.1	0.1	6.7	3.87	0.1	0.1	46.8	11.35

Numerical analysis

Preliminary data screening and transformation

All numerical analyses were performed using the programme CANOCO version 3.12 (ter Braak, 1988) and graphical representation of the data was made using the programme CALIBRATE (Juggins & ter Braak, 1992).

All environmental variables except pH were log(x+1) transformed to reduce the positive skewness. Diatom counts were transformed into percentages and all taxa with an abundance >1% and occurrence in at least three lakes were retained for further numerical analysis. Ninety-nine taxa out of a total of 202 diatom species found in the training set fulfilled the above criteria. Appendix 3 lists the retained taxa together with their number of occurrences, their effective occurrences and their codes used in Figures.

Environmental data were screened before numerical analysis in order to eliminate redundant and superfluous variables, unusual or 'outlier' samples and environmental variables that do not influence diatom distribution. The deletion of superfluous environmental data was especially essential for further analysis in Canonical Correspondence Analysis (CCA) because the number of environmental variables was equal to the number of samples minus one and this causes serious numerical problems (ter Braak, 1987a).

The screening of the environmental data followed Pienitz *et al.*, (1995) and ter Braak (1987a). Two criteria were imposed to identify multicolinear and superfluous variables:

- high variance inflation factors of the variables in an exploratory Detrended Correspondence Analysis (DCA) of the total environmental variables with all environmental variables regressed onto the DCA axes (ter Braak, 1988a);

- high correlation of the variables with the second 'arched' axes in the CCA (ter Braak, 1987a).

Four variables (COL, Ntot, NH_4 and COD) fulfilled both criteria and therefore were excluded from further analysis.

Outlier samples, i.e. samples having both unusual diatom composition and environmental characteristics were identified using exploratory Principal Component Analysis (PCA) with explanatory variables (environmental data) and DCA with response variables (diatom data). Site KOLA 17 obtained the highest score in PCA because it has unusually low water concentrations of major ions, high dissolved organic content and low pH (5.1). This site was also identified as a DOC outlier because of its diatom composition featuring relatively low species diversity (22 taxa), the highest abundance of *Stauroneis anceps* within the training set and dominance of coarse *Pinnularia* taxa.

Two outlier samples (KOLA8 and KOLA11) with extreme (> 8x) influence (ter Braak, 1990) were detected using the leverage diagnostics in the exploratory CCA. These sites have the highest lake area values. However, the exclusion of the above sites from the further CCA analysis has not improved the ordination and therefore finally these sites were retained.

Principal Component Analysis

PCA is an ordination technique for an indirect gradient analysis of variation in datasets. In PCA linear regression relationships between the variables and the underlying latent variables are assumed. PCA is appropriate for analysing the data with short (<2SD) gradients.

PCA was applied to study the major patterns of variation in the environmental data and to examine the relationship among environmental variables (ter Braak, 1988b). As the study focuses on the relationships between environmental variables rather than on relationships between sites. PCA on a correlation matrix of 20 response variables was performed (ter Braak, 1994).

The eigenvalues and cumulative variance of the data accounted by the first four PCA axes are shown in Table 3.

Table 3 Summary of eigenvalues and cumulative percentage variance of the PCA for 25 samples and 20 environmental variables

Axes	1	2	3	4
Eigenvalues	**0.249**	**0.196**	**0.163**	**0.088**
Cumulative variance	**24.9**	**44.5**	**60.8**	**69.6**

The first two PCA axes account for less than 50% variance in the environmental data and the eigenvalue of the third axes is comparable with the eigenvalue of the second axis. Therefore the third axes is also important for understanding the relationships between environmental data. The difference between the second and the third axes is relatively small suggesting that there might be some instability of the ordination (ter Braak, 1994). However, the exploratory PCA performed with the whole set (24) of environmental variables appeared to be even less stable - the difference between the second and the third axes was smaller and the iteration report in CANOCO revealed non-convergence between the first and second axes.

Figure 5 displays the PCA correlation biplots of Axis 1 against Axis 2. Environmental gradients are shown in (1). Arrangements of all 24 sites along the axes are shown in (2).

In PCA biplots the length of the arrows reflect the variance of the data and the arrow points towards the increase of the fitted value of a variable. Arrows that point in the same direction are positively correlated and the smaller the angle between them then the higher is the correlation. If the arrows are perpendicular there is no correlation and arrows pointing in opposite directions indicate negative correlation (ter Braak, 1987c).

Conductivity, sodium, chloride, magnesium, latitude and longitude are strongly ($P \leq 0.05$) negatively correlated with the first PCA axis; pH, alkalinity, calcium and potassium show high correlations with the second PCA axis. Physical variables (altitude, lake depth) are

correlated with the third axis which was revealed by examining the CANOCO output file.

Alkalinity, pH, potassium, calcium, silica and TOC are strongly intercorrelated. Sodium, chloride and sulphate are, as expected, related to latitude and longitude. Phosphorus shows no positive correlation with any of water chemistry parameters and it is negatively correlated with calcium-related group of variables (i.e. pH, alkalinity etc).

The PCA biplots illustrate the separation of tundra sites from sites with arboreal vegetation in the catchment. Forest sites have a wider range of chemical characteristics and this agrees with results of other studies (e.g. Pienitz & Smol, 1993). Alpine sites show more similarities to the forest and forest-tundra lake water chemistry than to tundra lakes and this contradicts the findings of other authors (e.g. Pienitz *et al*, 1995). This suggests that the chemistry of the lakes within this training set is influenced more by regional geology and sea ions rather than by catchment soils and vegetation. Tundra sites are arranged on the lower left-hand quadrant of the biplot, indicating the influence of seasalts on their chemistry and generally have a smaller range of chemical characteristics.

Detrended correspondence analysis

DCA is a technique for indirect gradient analysis which assumes a unimodal response of variables to the underlying gradients. DCA is usually applied to analyse the variation in floristic composition because it provides good results while ordinating data with large number of taxa, many zero values (i.e. for samples where species are absent) and long (>2 SD) gradients in floristic composition (Hill & Gauch, 1990).

Detrending by segments DCA was performed on the species matrix of 99 diatom taxa as recommended in (ter Braak, 1987c). Since the DCA is used to reveal ecological differences between species it is sensitive to rare species and therefore rare taxa were downweighted. The DCA axes were related by regression to external environmental variables which helps to interpret the resulting biplot.

The interpretation of the DCA is as follows: the distance between sites reflects the degree of the dissimilarity in species composition and the expected abundance of species at a site corresponds with the distance between the species position on the plot and the site position (Hill & Gauch, 1980).

The eigenvalues, cumulative percentage variance and the length of gradient accounted for the first 4 axes in DCA of the 99 diatom species with the 20 environmental variables regressed on to the axes are given in Table 4. Axis one and two account for 25.4% of the variance in the species data. The low percentage of variance explained is typical for data with many taxa and many zero values (Stevenson *et al.*, 1991) and occurs in many ecological studies (e.g. Pienitz *et al.*, 1995). The lengths of gradient for the first two DCA axes are 4.3 and 2.3 SDs respectively. This indicates that the sites arranged on the opposite ends of the diagram have hardly any species in common (Hill & Gauch, 1980).

No environmental variables have a statistically significant relationship with the DCA axes (*t*-values are no greater than 1.7). The most important, albeit not statistically significant variables, are pH for axis 1 and maximum depth for axis 2. Conductivity along with related

ions (e.g. sodium and potassium) also have some influence on the first DCA axis.

Table 4 **Summary of eigenvalues, cumulative percentage variance and the length of gradient of the DCA for 25 samples, 99 diatom species**

AXES	1	2	3	4	Total inertia
Eigenvalues	0.411	0.317	0.226	0.148	2.862
Length of gradient	4.302	2.293	2.174	1.194	
Cumulative % variance of the species data	14.4	25.4	33.3	38.5	

DCA revealed the major patterns in diatom distribution within the training set which are presented in Figure 6. Diagram (2) shows variation in species data, diagram (1) illustrates the arrangement of sites. Full names of diatom taxa together with their codes used in the Figures are given in Appendix 3.

Since pH is negatively correlated with axis 1, the most acidified, dilute and clear lakes (e.g. KOLA17, KOLA19 and KOLA3) are grouped on the right side of the biplot (Figure 6). Therefore, associated acidophilic species such as *Navicula hoefleri, N. mediocris, N. subtilissima, N. tenuicephala, N.* cf. *impexa* (SWAP), *Frustulia rhomboides, F. rhomboides* var. *saxonica*, *Stauroneis anceps* var. *anceps*, *Aulacoseira distans* var. *nivalis*, A. *perglabra* var. *perglabra, Cymbella perpusilla, C. naviculiformis, Eunotia exigua, E. serra, Pinnularia mesolepta, P. biceps,* and *P. stomatophora* also occur on the right part of the biplot.

The left-hand side of the plot is occupied by species occurring in more alkaline waters with higher conductivity and cation concentrations, predominantly small benthic *Achnanthes* and *Fragilaria: Achnanthes didyma, A. suchlandtii, A. nodosa, A. pseudoswazi, A.curtissima, A. levanderi, A. flexella, Cyclotella pseudostelligera, Fragilaria elliptica, F. construens* var. *venter, F.* cf. *brevistriata* and *F. pseudoconstruens.*

The lower left quadrant of the biplot (where the shallow sites are grouped) is also occupied for the most part by the complex of benthic *Fragilaria*, *Achnanthes* and *Aulacoseira*: *Fragilaria pinnata*, *F. virescens* var. *exigua, F. bicapitata, Achnanthes detha, A. kriegeri, Aulacoseira lirata* var. *lirata, A. italica* var. *subarctica, A. nygardii, A. italica* var. *valida. Nitzschia palea* and *N.* cf. *gracilis* (SWAP) also occur on the lower left quadrant of the biplot.

The upper half of the biplot is dominated by the complex of deep water centric diatoms: *Cyclotella rossi, C. pseudostelligera, Aulacoseira lirata* var. *subarctica, A. distans* var. *nivalis, A. lirata* var. *alpigena, A. perglabra* var. *florinae*. Several *Navicula* species i.e. *N. cocconeiformis, N. digitulus, N. pseudoscutiformis* together with *Stauroneis anceps* var. *gracilis* and *Surirella linearis* also have high scores on axis 2.

High scores and therefore extreme positions of several species (which are not abundant in the whole training set) close to the edge of the diagram (e.g. *Stauroneis anceps* var. *anceps*, *Navicula digitulus,* and *Achnanthes suchlandtii*) indicate that these particular species happen to occur in lakes with contrasting diatom assemblages and such species can not be considered as indicators of acid conditions (*S. anceps*), alkaline conditions (*A. suchlandtii*) or deep water alpine environments (*N. digitulis*). Although the rare species were downweighted they still appeared to influence the analysis which is one of the shortcomings of DCA (ter Braak, 1987c).

Hence, the DCA effectively separates diatom species of the training set according to their water acidity (pH) and microhabitat preferences (maximum depth). The influence of these environmental variables on the diatom distribution have been found in many ecological studies (e.g. Smol, 1988; Dixit *et al.*, 1991; Stevenson *et al.*, 1991; Pienitz *et al.,* 1995a).

Since forest lakes have the longest pH gradient they are arranged along the first axes according to their pH (Figure 6). Forest/tundra lakes are predominantly grouped on the left-hand quadrant of the plot except for the site KOLA5 which is the deepest lake within the dataset and is therefore positioned on the upper part of the biplot. Tundra lakes are grouped on the left upper part of the biplot and in the centre since they are generally deeper and more alkaline than most of the forest lakes.

Canonical Correspondence Analysis

CCA is the unimodal method of direct ordination. CCA allows investigation of the direct relationship between floristic composition and environmental variables, because in contrast to indirect techniques it uses both species and environmental data to arrange sites along the ordination axes which are constrained to be the function of environment (ter Braak, 1995).

CCAs were performed on the species matrix of 99 taxa and 20 environmental variables to identify the explanatory variables directly accounting for the variation in floristic composition. In all CCAs rare species were downweighted. The species scores were scaled to be weighted averages of the site scores which is the most popular type of scaling used in similar palaeoecological studies (e.g. Pienitz & Smol, 1993; Pienitz *et al.*, 1995). Sites are plotted as linear combinations of environmental variables (ter Braak, 1987c).

Forward selection in CCA was used to determine the minimal set of environmental variables accounting for the significant proportion in the species variation. Canonical coefficients, their *t*-values and inter-set coefficients were used to assess the relative importance of the selected environmental variables to the ordination axes.

Partial CCAs and variance partitioning were performed to assess the proportion of species variance explained by different sets of environmental variables and individual environmental determinants (ter Braak, 1988b).

Monte Carlo permutation tests were performed to test the significance of canonical axes and variables. A Bonferonni correction (Manly, 1992) was also used to establish an appropriate value of P from which successive variables were retained as being statistically significant (P

≤ 0.05 for the first selected variable, $P \leq 0.025$ for the second selected variable and $P \leq 0.016$ for the third variable). Monte Carlo tests with 99 unrestricted permutations were used to establish the statistical significance of the axes. Monte Carlo tests with 999 unrestricted permutations were applied to assess the statistical significance to the third decimal place.

CCA1 of the 20 environmental variables and 99 diatom taxa

The first two CCA axes ($\lambda_1 = 0.393$ and $\lambda_2 = 0.316$) explain a total 23.6% of the variance in the diatom data species variance which is similar to the results of the DCA (25.4%). The species-environmental correlations are high for the first axes (0.992 and 0.990). Twenty environmental variables account for 27.2 % in species variance which is rather low compared to other studies (e.g. Vyverman & Sabbe, 1995; Pienitz, 1995). However, Monte Carlo permutation tests of significance of axis 1 with 99 unrestricted permutations showed that the axis is not statistically significant ($P = 0.62$). This means that the variance of the species data is not sufficiently explained by the set of environmental variables. The high species-environmental correlations in this case do not mean a strong species-environmental relationship (ter Braak, 1995). Several reasons could lead to the insignificant CCA ordination:

1. superfluous and multicolinear environmental variables;
2. presence of outlier sites or sites with unusual environmental variables characteristics which bias the environmental gradients;
3. short environmental gradients.

Subsequent analyses were undertaken in an attempt to evaluate the above hypotheses. Preliminary data screening was performed in order to identify and eliminate the superfluous variables. Four variables which showed the highest variance inflation factors and had no ecological importance in this study were excluded. However, even after exclusion of these variables several environmental variables (for the most part major ions and conductivity) still remained multicolinear.

CCA2 of three forward selected variables, 25 samples and 99 diatom taxa

Separate CCA3 was performed with the three forward selected statistically significant variables - pH, Long and Max D. Table 5 gives the summary statistics for the CCA2. The first canonical axis is significant ($P = 0.01$) and it explains 42.5% of variance in the diatom-environment relationship which is comparable with other studies (e.g. Pienitz & Smol, 1993). Percentage of explained variance in the diatom taxa (9.6) is similar to the CCA1 of the full environmental matrix which means that the three selected environmental variables contribute substantially to the ordination. However, the second CCA axis is not significant, ($P = 0.13$) and the ratio of λ_2/λ_3 is low.

Table 6 presents the canonical coefficients, their *t*-values and intra-set correlations of the variables with the CCA axes. Longitude and pH strongly correlate with the first significant axis and maximum depth shows a strong relationship with the second axis. The CCA biplot is presented in Figure 7. In principle, the results of forward selection are similar to the CCA2 and DCA and this suggests that the three selected parameters are the most important explanatory variables in the training set.

Table 5 **CCA2: summary statistics for three forward selected environmental variables, 25 samples and 99 species**

AXES	1	2	3	4	Total inertia
Eigenvalues	0.289	0.206	0.185	0.322	3.002
Cumulative % variance of species data	9.6	16.5	22.7	33.4	
Cumulative % variance of the species-environment correlation	42.5	72.8	100.0	-	
Sum of canonical eigenvalues					0.680

Table 6 **CCA2: Canonical coefficients of the three forward selected variables, their *t*-values and intra-set correlations with axes. All *t*-values are significant at $P \leq 0.05$**

Environmental variable	Canonical coefficients		t-values of canonical coefficients		Intra-set correlations	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
pH	-0.842	-0.286	-6.72	-2.04	-0.547	-0.095
Longitude	-0.745	0.4396	-6.18	3.25	-0.520	0.4546
Maximum depth	0.364	0.8850	2.94	6.36	0.0812	0.7083

CCA3 with forward selection of the environmental dataset with longitude excluded, 25 samples and 99 diatom taxa

The fact that longitude is chosen in forward selection means that it represents a major environmental gradient. Basic statistics and PCA of environmental data revealed that longitude is correlated with major ions and conductivity. Due to the limited geographical range it is unlikely that longitude represents the climatic or vegetational gradient and the above CCAs have also revealed no clear geographical patterns in the arrangement of sites in relation to the longitude vector. Thus, longitude may be a substitute for another gradient, most probably for conductivity and/or major ions.

The following analysis was conducted to test this hypothesis: longitude was excluded from the CCA3 and the following variables were forward selected (in descending order): pH, maximum depth and sodium. Summary statistics and canonical coefficients together with their *t*-values and intra-set correlations (Tables 7 and 8) agree closely with the respective values of CCA2 (Tables 5 and 6). Eigenvalues of CCA3 have slightly lower values and the first axis

accounts for less variance both in the species data and in species-environmental correlation. However, these differences are not large. The first axis also shows a strong relationship with pH and sodium, the second axis correlates well with maximum depth.

A Monte Carlo permutation test revealed the first axis ($P = 0.02$) to be statistically significant and the second axis not to be significant. Figure 8 shows the CCA3 biplot which is almost identical to the biplot of the CCA2.

Thus it appears that longitude in this training set represents the gradient of major ions, most probably of seasalts.

Table 7 **CCA3: summary statistics for three forward selected environmental variables, 25 samples and 99 species**

AXES	1	2	3	4	Total inertia
Eigenvalues	0.260	0.214	0.169	0.330	3.002
Cumulative % variance of species data	8.7	15.8	21.4	33.4	
Cumulative % variance of the species-environment correlation	40.4	73.7	100.0	-	
Sum of canonical eigenvalues					0.642

Table 8 **CCA3: Canonical coefficients of the three forward selected variables, their *t*-values and intra-set correlations with axes. All *t*-values are significant at $P \leq 0.05$**

Environmental variable	Canonical coefficients		t-values of canonical coefficients		Intra-set correlations	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
pH	-0.772	-0.374	-5.04	-2.63	-0.657	-0.04
Sodium	-0.567	0.466	-3.85	3.40	-0.564	0.346
Maximum depth	0.228	0.943	1.50	6.69	0.018	0.712

CCA4 with forward selection of 23 samples, 20 environmental variables and 99 diatom taxa

As suggested above, the reason for the insignificance in CCA ordination may be the influence of the outlier sites or sites with unusual parameters. Outliers were discussed above and their removal has shown no improvement in ordination. Nonetheless, two alpine sites (KOLA2 and KOLA25) which albeit were not identified as outliers in preliminary data screening, could bias the ordination. There is no distinct altitude gradient in the data and the presence of two high altitude tundra and forest/tundra sites within the forest zone might affect at least the forest and peatland gradient. The water chemistry of KOLA2 and KOLA25 fit better the chemistry of forested lakes than the chemistry of tundra lakes. The exception is the concentration of iron which is lower in these two lakes compared to other lakes of the forest zone but the difference is not great.

CCA4 was performed to test whether the presence of KOLA2 and KOLA25 bias the ordination. The resultant full CCA on the whole matrix of environmental variables shows no better results and the first axis remains insignificant. Summary statistics are displayed in Table 9. The results are close to CCA1 though the first eigenvalue in CCA4 is slightly higher and the proportion of explained variance both in species data and in species-environment relationships is also higher.

The following variables were chosen by forward selection in CCA4 (in descending order of potential variance explained): conductivity, maximum depth and pH. Other variables selected were statistically insignificant and therefore they were omitted. The summary statistics of the CCA performed with the above variables are presented in Table 10. This analysis has all the shortcomings of the previous analyses: the insignificance of the second axis while the first axis is significant, and the low ratio of second eigenvalue to the third eigenvalue.

This CCA is similar to the previous analyses in terms of correlations of variables with axes presented in Table 11. Figure 9 shows the correlation biplot of CCA4. Conductivity and pH contribute to the first significant axis, maximum depth respectively to the second axis. The only difference is that conductivity is the variable which shows the stronger correlation with the axis other than pH.

CCA4 confirmed the previous suggestions that longitude influences diatom distribution in the same way as sodium or conductivity. Therefore, longitude apparently can represent the distance from the sea rather than climate or catchment vegetation. Longitude may still have influence independent from ions or conductivity and this can be tested with variance partitioning.

Table 9 CCA4: summary statistics for 23 samples, 20 environmental variables and 99 species

AXES	1	2	3	4	Total inertia
Eigenvalues	0.401	0.345	0.309	0.237	2.807
Cumulative % variance of species data	14.3	26.6	37.6	46.1	
Cumulative % variance of the species-environment correlation	15.5	28.9	40.8	50.0	
Sum of canonical eigenvalues					2.587

Table 10 CCA4: summary statistics for three forward selected environmental variables, 23 samples and 99 species

AXES	1	2	3	4	Total inertia
Eigenvalues	0.266	0.234	0.185	0.339	2.807
Cumulative % variance of species data	9.5	17.8	34.4	36.5	
Cumulative % variance of the species-environment correlation	38.8	73.0	100.0	-	
Sum of canonical eigenvalues					0.682

Table 11 CCA4: Canonical coefficients of the three forward selected variables, their *t*-values and intra-set correlations with axes. All *t*-values are significant at $P \leq 0.05$

Environmental variable	Canonical coefficients		t-values of canonical coefficients		Intra-set correlations	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
pH	-0.3819	-0.5412	-2.34	-3.829	-0.539	-0.163
Conductivity	-0.8105	0.2769	-5.14	2.02	-0.787	0.126
Maximum depth	0.0527	0.9926	0.34	7.45	0.074	0.754

Hypothesis testing and partitioning the variance

To test the hypothesis whether longitude has any effect on diatom distribution independent from seasalts (sodium and chloride) variance partitioning was performed following Borcard *et al.*, (1992).

Table 12 shows the results of partitioning the variance. The following analyses were performed to identify the percentage of variance explained by the seasalts and longitude:

1 to identify the unexplained variance: CCA6.1 of the species matrix constrained by sodium, chloride and longitude. Unexplained variance = 100% - explained variance in CCA6.1;
2 to identify the total variance explained by sodium and chloride: CCA6.2 of the species matrix constrained by sodium and chloride;
3 to identify the variance explained by sodium and chloride independent of longitude: CCA6.3 of the species matrix constrained by Na and Cl and with longitude as a covariable;
4 to identify the total variance explained by longitude: CCA6.4 of species matrix constrained by longitude;
5 variance explained together by covariance of longitude, sodium and chloride: step (2) - (3);
6 to identify variance explained by longitude independent of sodium and chloride: CCA6.5. of species matrix constrained by longitude and sodium and chloride as variables. However, the first axis in this CCA was insignificant, so the independent contribution of longitude was assessed as (4) - (5).

The results show that 36% of the total variance explained by longitude covaries with the seasalts and 64% of the total variance explained by longitude is independent of the influence of seasalts.

All previous estimations conclude that pH makes the strongest and most significant contribution to the variance in the species data and therefore a pH-inference model could be developed. To assess directly the influence of pH on the variation in the floristic composition a series of partial CCAs and partitioning of variance were performed (ter Braak, 1988a, Borcard *et al.*, 1992).

CCA7.1 - CCA7.4 were carried out with the diatom matrix constrained by pH as the sole environmental variable (Table 13) and it was analogous to the procedure described above. Unexplained variance = 100% - species variance of CCA1 with full environmental dataset.

The total variance in diatom taxa explained by pH is 7.5% and this result is statistically significant. However, the eigenvalue of the second unconstrained axis in the CCA7.1 of pH as the only environmental variable is greater than the eigenvalue of the first pH-constrained axis. This suggests that the pH gradient is too short for the successful development of a pH transfer function (ter Braak, 1988a).

The variance explained by pH independent of physical variables i.e. latitude, longitude, altitude, maximum depth, lake area, percentage forest and percentage peat is 6.2% and this

comprises 82.6% of the total variance explained by pH. Therefore, pH covaries a little with measured physical characteristics.

All other partial CCAs were assessed as statistically insignificant and they do not merit interpretation.

Partial CCA7.1 - 7.4 showed that pH is independent of physical variables and gives a statistically significant gradient in environmental data. However, the pH gradient in this training set is inadequate to explain much variation in diatom data and therefore the training set requires amalgamation with other data.

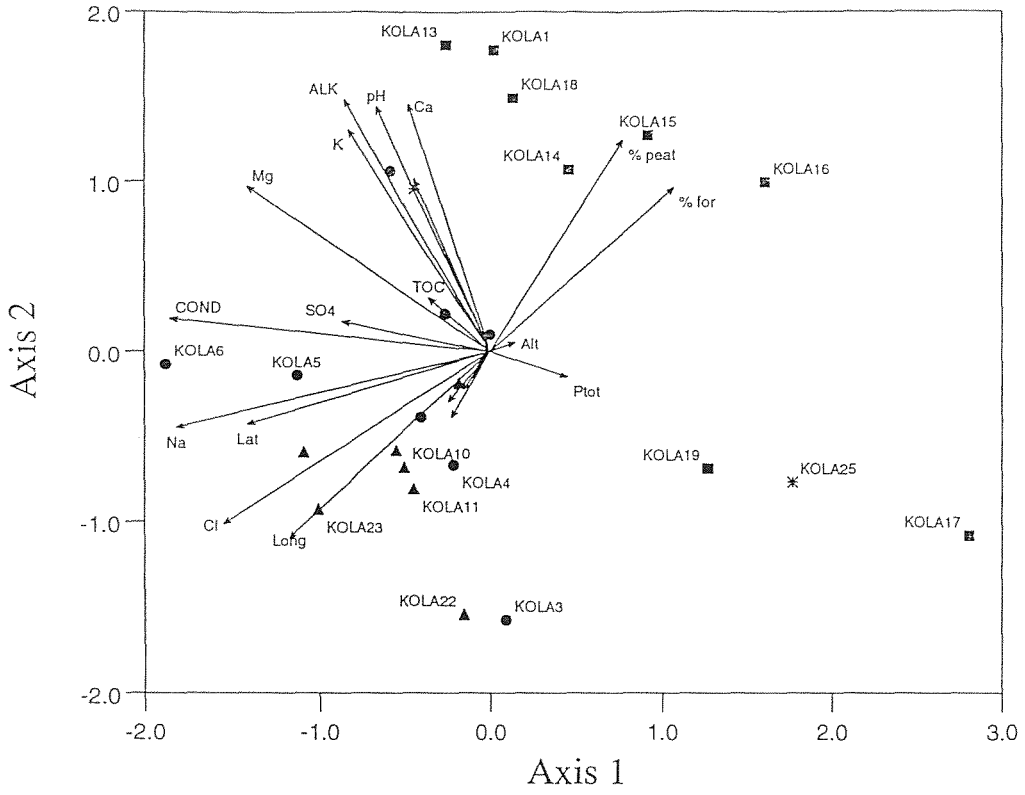
Table 12 **Results of partitioning the variance in the diatom matrix of 99 taxa. Entries are percentages of the total variance in the data set. *P* = probability in Monte Carlo tests of significance of the first axis**

Source of variance		Percentage	*P*
1.	Unexplained variance	81.3	0.07
2.	Variance explained by seasalts independent of longitude	12.4	0.05
3.	Variance explained by longitude independent of seasalts	4.9	0.04
4.	Variance explained by longitude covarying with seasalts	2.7	0.05
5.	Variance explained by seasalts as sole environmental variables	15.1	0.04
6.	Variance explained by longitude as sole environmental variable	7.6	0.01

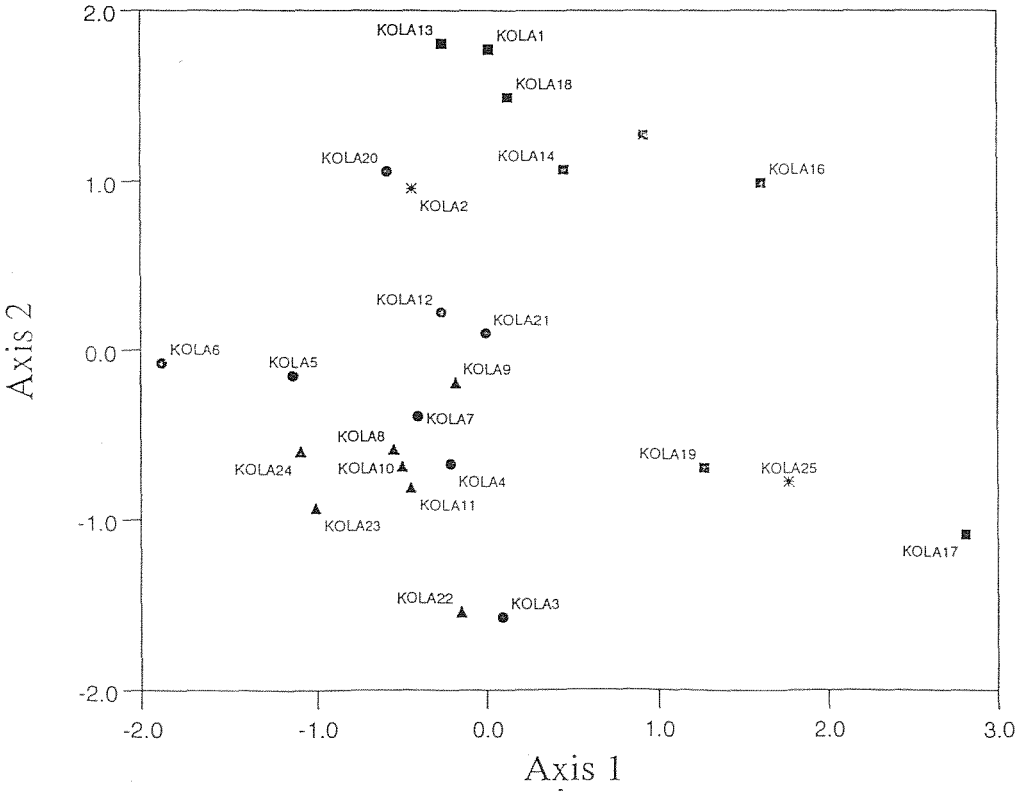
Table 13 **Results of partitioning the variance in the partial CCA7.1 - CCA7.4 of diatom matrix of 99 taxa and pH as the sole explanatory variable. Entries are percentages of the total variance in the data set. *P* = probability in Monte Carlo tests of significance of the first axis**

Source of variance		Percentage	*P*
1.	Unexplained variance	58.6	0.62
2.	Variance explained by pH as the sole environmental variable	7.5	0.02
3.	Variance explained by pH independent of all other environmental variables	4.0	0.37
4.	Variance explained by pH independent of water chemistry variables	4.5	0.20
5.	Variance explained by pH independent of physical variables	8.8	0.01
6.	Variance explained by pH covarying with physical variables	6.8	0.01

Figure 5 PCA biplots for 25 site scores (1), and 20 environmental variable scores (2). Forest sites are shown as filled squares, forest - tundra sites as filled triangles, alpine sites as stars

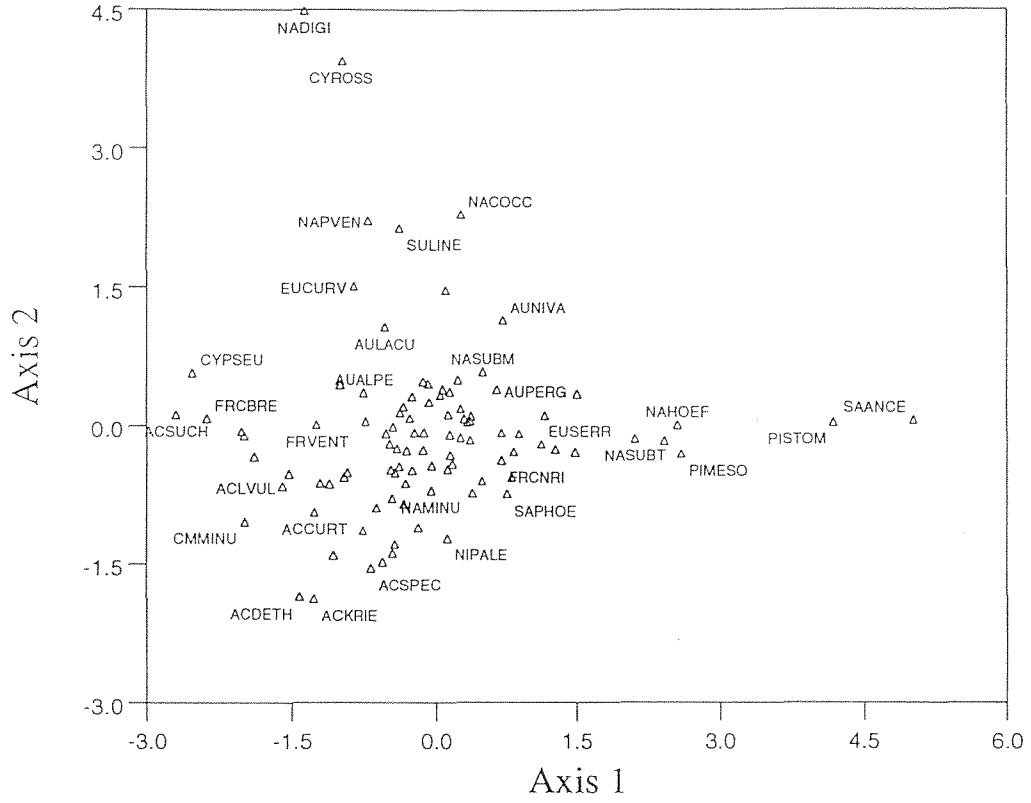


(2)

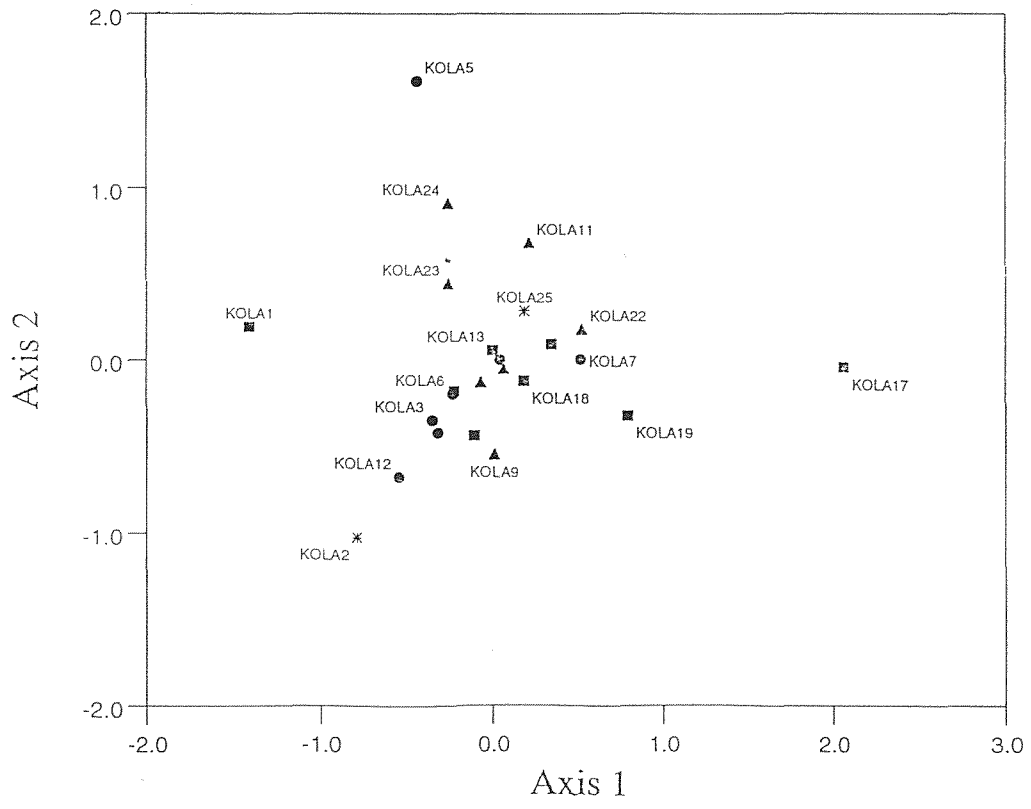


(1)

Figure 6 **DCA plots of 25 site scores (1) and 99 species scores (2). Species centroides are shown as open triangles. Full diatom names are given in Appendix 3**



(2)



(1)

Figure 7 **CCA2 biplot of 25 site scores and three forward selected environmental variables**

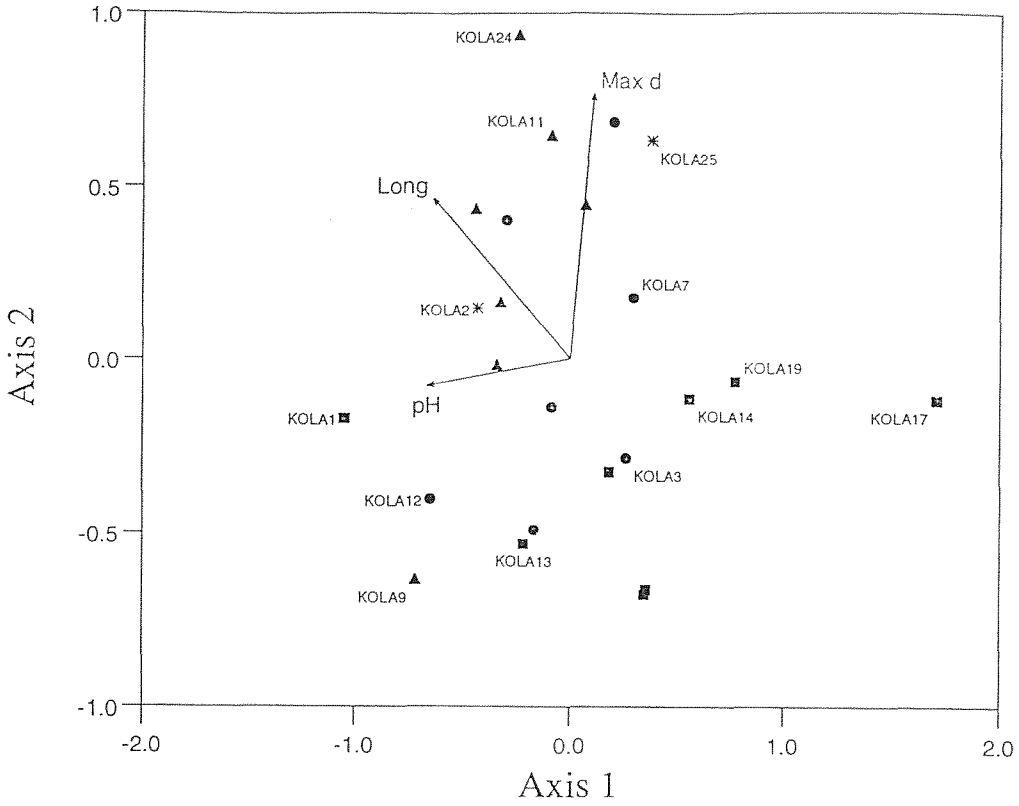


Figure 8 **CCA3 biplot of 25 site scores and three environmental variables forward selected in the absence of longitude**

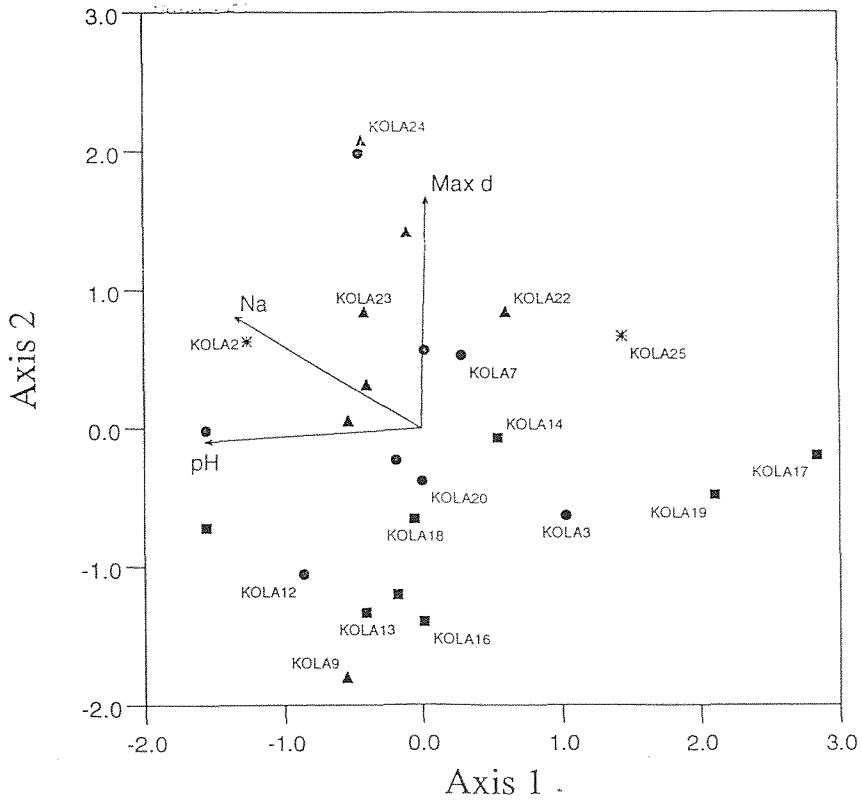
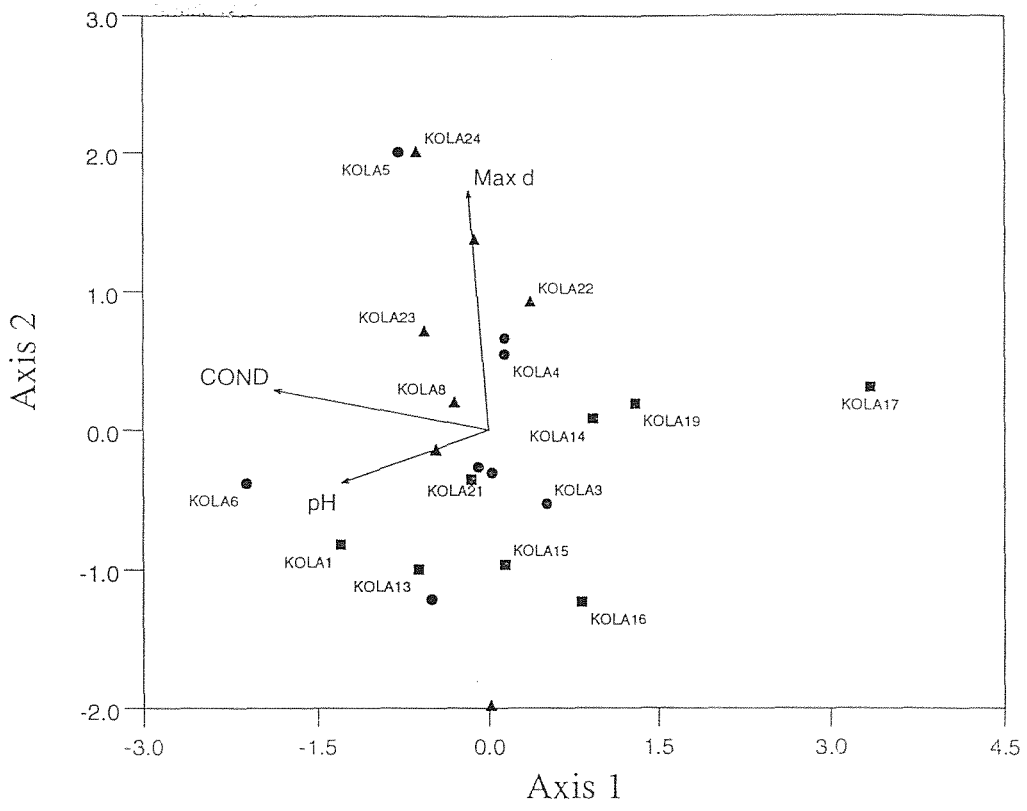


Figure 9 **CCA4 biplot of 23 site scores and three forward selected environmental variables**



A trial pH transfer function

As shown in the above sections, the pH gradient is inadequate for the successful generation of the inference model and more data are required for a statistically robust model. Nonetheless, weighted-averaging (WA) regression and calibration were performed in order to develop a pH-transfer function with the current data and to estimate diatom pH optima and tolerances in the current training set. This information will help with the selection of lakes for amalgamating data into this training set.

WA methods are currently well established (e.g. Birks *et al.*, 1990a,b; ter Braak & Juggins, 1993) and they have been successfully used for inferring different chemical and physical parameters. The assumption made in WA techniques is that diatom taxa are most abundant in the lakes with pH (or other parameter of interest influencing diatom distribution) close to their ecological optima (Birks *et al.*, 1990a). The optimum or 'indicator value' (ter Braak, 1987b) for each taxon is estimated as the average of lake pH where the taxon occurs weighted by its relative abundance. Taxa that are absent are given zero weight.

Quantitative estimates of pH changes in the past can then be made with WA calibration using the diatom pH optima estimates and tolerance and their relative abundances in the sediment core.

The programme CALIBRATE (Juggins & ter Braak, 1992) was used to perform WA regression on the training set (25 samples, 99 diatom taxa) to derive and compare diatom pH optima and tolerances using simple WA and tolerance-downweighted WA.

Inverse deshrinking was used to correct the effect of shrinkage of inferred values, because averages are taken twice (ter Braak & van Dam, 1989). The inverse deshrinking was chosen because it provides more accurate results across the environmental gradient as a whole compared to classical deshrinking which works better with extreme values (ter Braak & van Dam, 1989).

The predictable ability of the transfer function was assessed by the correlation coefficient (r^2), the apparent root mean squared error (RMSE) of prediction, jackknifed error of prediction, and the distribution of residual (observed - predicted) values of lake pH. Jackknifing involves a pH prediction for each sample based on the remainder of samples in the dataset other than the one not to be tested (Hinkley, 1983).

Summary statistics are given in Table 14. The apparent RMSE is rather high but comparable with other authors (e.g. Stevenson *et al.*, 1991). However, RMSE is not always a reliable measure, as it is usually underestimated because of the re-substitution of samples during its computation. The jackknifed cross-validation error has greater value and is more realistic. The correlation values generally agree with other relevant studies (e.g. Dixit *et al.*, 1991).

Figure 10 illustrates the predictive abilities of the transfer function. The following diagrams are plotted: (1) observed pH of the lakes against predicted WA pH; (2) observed pH against predicted WA-tol pH; (3) observed pH against WA residuals (observed - inferred WA pH) and (4) observed pH against WA-tol residuals. The scatter plots show that the predicted pH values fit rather well with the observed pH values. Two outliers (sites KOLA3 and KOLA19) have the highest residuals both in WA and WA-tol regression and calibration. WA and WA-tolerance regression show similar results.

The WA pH optima of 99 diatom taxa, their tolerances, number of diatom occurrences and their effective number of occurrences (N2) (Hill, 1973) are shown in Appendix 3. N2 is a measure of the diversity reflecting both taxon abundance and occurrence, so that taxa occurring in many sites with low abundance will have a low N2.

On the whole, the estimated diatom optima agree with estimated optima in other studies i.e. common acid-tolerant taxa have lower optima than common alkaline-tolerant taxa (e.g. Dixit *et al.*, 1991; Schmidt & Psenner, 1992). For example, typical acidophilous diatoms *Navicula hoefleri, N. subtilissima, Pinnularia biceps, P. microstauron, Frustulia rhomboides, F. rhomboides* var. *saxonica* and *Stauroneis anceps* obtained relatively low WA optima and this agrees with other authors (e.g. Stevenson *et al.*, 1991). Most of the small *Fragilaria* taxa obtained relatively high optima and this is also in agreement with the estimations made in other works (e.g. Schmidt & Psenner, 1992).

However, there are some disagreements between optima obtained in this study and results of other studies. For instance, *Achnanthes marginulata* estimated as an acid-tolerant species with optima ranging from pH 5.2 to 5.86 (Stevenson *et al.*, 1991; Dixit *et al.*, 1991; Schmidt &

Psenner 1992) was assigned an optima of 6.51 by the current model. *Fragilaria pinnata* has an unusually low optima of 5.82 in this training set and other authors (Stevenson *et al.*, 1991) estimate its optima as 6.3. The optima of *Fragilaria virescens* is largely overestimated in this study (6.29) compared to Dixit *et al.* (1991) (5.79) and Stevenson *et al.* (1991) (5.6) but it is close to the pH optima derived by Schmidt & Psenner (1992) (6.2).

On the whole, the current model tends to overestimate diatom optima even in comparison with training sets of lakes with similar pH.

The following criteria were used for preliminary identification of indicator species (Birks *et al.*, 1990; Stevenson *et al.*, 1991):

- more than 10% of a taxon variance should be explained by pH (in a CCA with pH as the sole environmental variable);
- the WA tolerance should be less than the average tolerance for the training set (0.54);
- a taxon should occur in more than five lakes.

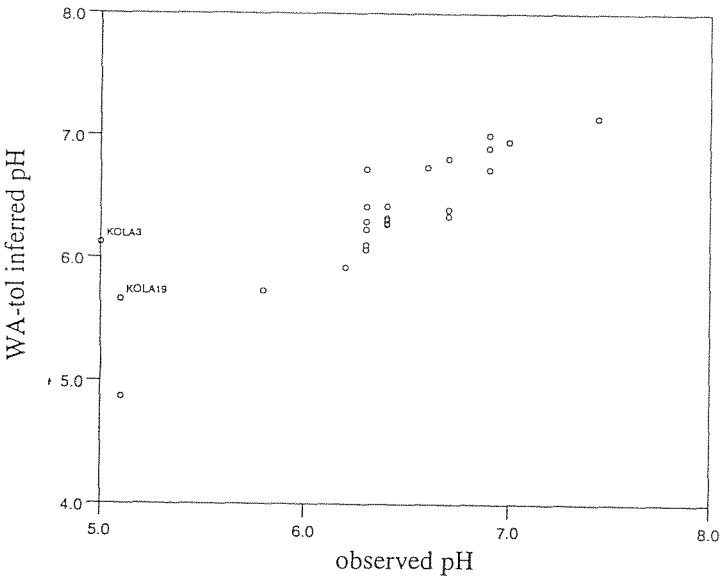
Only three species (*Achnanthes curtissima, Cymbella cesatii* and *Cymbella minuta*) fulfil the above criteria. Although the pH effectively explains more than 10% of variance of a high number of diatom taxa, the tolerance of many species was too high and therefore they can not be used as reliable indicators of pH changes. The indicator species obtained in other studies (Stevenson *et al.*, 1991) comprise different diatom species.

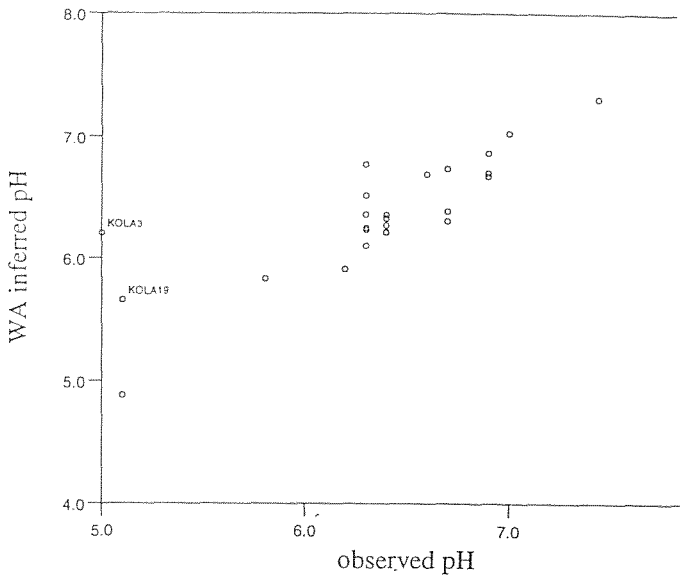
On the whole, the WA pH optima derived for this training set show partial agreement with the optima obtained in other works. The transfer function demonstrates good predictive abilities but the pH gradient in this training set is too short to obtain reliable indicator species.

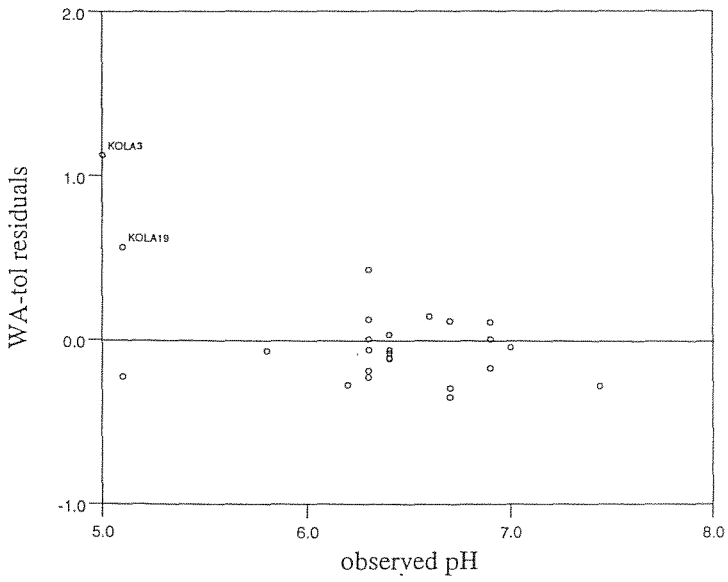
Table 14 Summary statistics for pH WA and tolerance-downweighted WA regression of the full training set (25 samples, 99 diatom taxa with abundance > 1% and occurrence in at least three lakes)

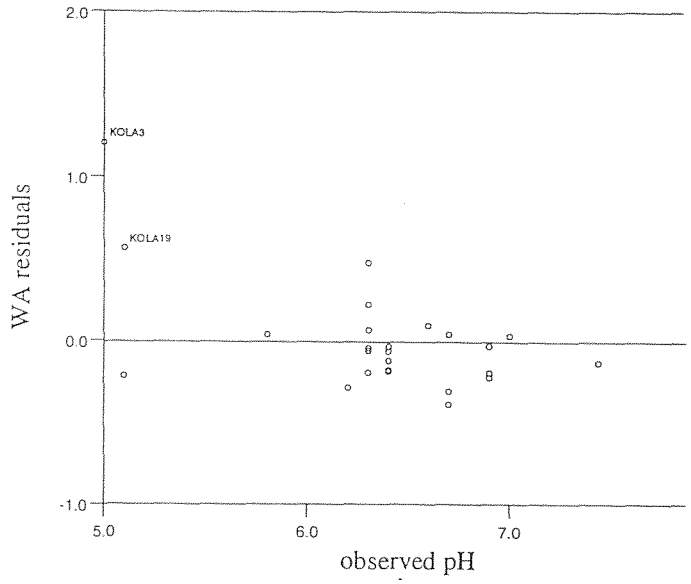
Type of WA regression	r^2	RMSE
Simple WA	0.6746	0.3277
Tolerance-downweighted WA	0.7070	0.3109
Jackknifed WA	0.0975	0.5501
Jackknifed tolerance-downweighted WA	0.1105	0.5458

Figure 10 **Plots of observed versus diatom-inferred pH and observed versus residual pH, based on regression and WA-tolerance regression and calibration**









3. GENERAL DISCUSSION

On the whole, the results of exploring relationships between diatom taxa and selected environmental variables indicate that pH, maximum depth and longitude are the variables explaining the major proportion of the variance in the species data. This is in agreement with the findings of other studies. The high influence of pH on diatom composition is a well known and established fact (e.g. Hustedt, 1937 - 1939; Nygaard, 1949; Battarbee, 1984) and many pH based transfer functions have been successfully developed and used for inferring past acidity changes (e.g Stevenson *et al.*, 1991; Dixit *et al.*, 1991). However, such studies usually involve larger training sets or training sets of lakes with steeper gradients. Clearly, the current training set requires enhancing with more data. Preferably, it could be extended to include lakes from the adjacent regions i.e. Northern Fennoscandia with a greater range of pH.

The other variable, which appeared important in this training set is maximum depth. It has also been recognized as an important factor accounting for distribution of diatom taxa in works of many authors (e.g. Hall & Smol, 1992; Pienitz & Smol, 1993; Smol, 1988). The significance of water depth is correlated with light penetration and consequently with water transparency. Therefore, the influence of water depth is greater in clear water oligotrophic lakes which are usually common in the tundra or alpine zones (Smol, 1988). The current results confirm this.

Longitude and latitude were included in this training set as cumulative characteristics of climate because the vegetation zones show both latitudinal and longitudinal zonation in the Kola Peninsula due to the peculiarities of the Gulf Stream. Longitude appeared to be a much stronger variable than latitude and this is due to the different ranges of these two variables. Longitude has an almost three times longer gradient than latitude. Longitude reflected the influence of the sea ions i.e. sodium and chloride. The Variance partitioning revealed that about one third of variation in floristic composition could be explained by longitude covariance with seasalts. The other two thirds of the variance remains undetermined. The unexplained part of variance controlled by longitude might be related to climatic characteristics. However, more data are required to test this hypothesis.

Results of the numerical analyses also suggest that the main cause of ordination instability is the lack of appropriate environmental gradients. Therefore the next stage of this project should be the selection of appropriate sites from other existing training sets because the time available in this project does not allow for collecting new samples from the study area. The most appropriate training set seems to be the **MOLAR** (Mountain Lakes Research) set of lakes comprising tundra-type alpine lakes from different mountain regions of Europe (Scandinavia, Alps, etc). Scandinavian lakes should provide the best fit with the current training set because of many similarities (geology, geomorphology, geochemistry, vegetation, climate etc.) between the Kola Peninsula and the whole Fenno-Scandinavian Peninsula. The recent survey on diatom and chrysophyte flora of Northern Fennoscandia (Pienitz, *et al.*, 1995) revealed many similar patterns of diatom composition between the Kola Peninsula and Northern Fennoscandia.

4. SUMMARY

Analyses of quality and range of environmental variables, basic statistics, numerical analyses of relationships between species and environmental matrices and analysis of the resultant transfer function permit the following conclusions:

- pH, maximum depth and longitude/sodium and conductivity are the most important explanatory variables in this training set;
- pH appeared to be the strongest explanatory variable in the training set, its effect on the diatom distribution is the most clear and its influence is statistically significant because pH in all analyses showed strong correlations with the first significant CCA axis. It accounts for variance in the diatom data independent of the physical variables;
- the pH gradient is still inadequate for the useful development of inference models and the training set requires amalgamation with additional data;
- longitude explains the variation in floristic composition in a similar way to sodium and conductivity; however, only one third of the variance explained by longitude covaries with the seasalts;
- the main reason for the lack of stability in the PCA and CCA ordinations is apparently short gradients of measured environmental variables because CCAs still remained unstable after excluding multicolinear and superfluous variables and after excluding the outlier sites and sites which can bias the environmental gradients.

BIBLIOGRAPHY

Atlas of Murmansk region, 1971. 33 pp. (in Russian).

Battarbee, R.W. 1984. Diatom analysis and the acidification of lakes. Phil. Trans. R.Soc. Lond., B **305**: 451-477.

Battarbee, R.W., 1986. Diatom analysis. In: B.E. Berglund (ed.), Handbook of Holocene Palaeoecology and Palaeohydrology. Wiley, Chichester: pp. 527-570.

Battarbee, R.W., 1992. Critical loads and acid deposition for UK freshwaters. ECRC UCL Research Paper No 5.

Battarbee, R.W., Anderson, N.J., Appleby, P.G., Flower, R.J., Fritz, S.C., Haworth, E.Y., Higgitt, S., Jones, V.J., Kreiser, A., Munro, M.A.R., Natkanski, J., Oldfield, F., Patrick, S.T., Richardson, N.G., Rippey, B. & Stevenson, A.C., 1988. Lake acidification in the United Kingdom 1800-1986. Evidence from analysis of lake sediments. ENSIS, London: 68 pp.

Birks, H.H., 1991. Holocene vegetational history and climatic change in west Spitsbergen - plant macrofossils from Skardtjorna, an Antarctic lake, The Holocene, **1,3**: 209-218.

Birks, H.J.B., 1986. Late-Quaternary biotic changes in terrestrial and lacustrine environments, with particular reference to north-west Europe. In: B.E. Berglund (ed.), Handbook of Holocene Palaeoecology and Palaeohydrology. Wiley, Chichester: pp. 3-66.

Birks, H.J.B. & Birks, H.H., 1980. Quaternary Palaeoecology, Botany School, University of Cambridge, 288 pp.

Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C. & ter Braak, C.J.F., 1990a. Diatoms and pH reconstruction. Phil. Trans. R. Soc., Lond. **B 327**: 263-278.

Birks, H.J.B., Juggins, S. & Line, J.M., 1990b. Lake surface-water chemistry reconstructions from palaeoecological data. In: B.J. Mason (ed.), The Surface Waters Acidification Programme. Cambridge University Press, Cambridge: pp 301-311.

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

Borland International, 1988. Paradox user's guide, release 3.0. Borland International, Scotts Valley, California: 310 pp.

Cameron, N.G., Rose, N.L., Appleby, P.G., Schnell, O.A., Battarbee, R.W., Patrick, S.T. & Flower, R.J., 1993. AL:PE 1: Palaeolimnology. UCL ECRC Research Report No. 4.

Cleve-Euler, A., 1951-1955. Die diatomeen von Schweden und Finnland. Kungl. Svenska Vet. Handlinger, Almqvist & Wiksell, Stockholm.

Cwynar, L.C & Spear, R.W., 1991. Reversion of forest to tundra in the central Yukon. Ecology, **72**: 202-212.

Davis, M.B. & Botkin, D.B., 1985. Sensitivity of cool-temperature forests and their fossil pollen record to rapid temperature change. Quatern. Res., **23**: 327-340.

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

Dixit, S.S., Dixit A.S.A, & Smol, J.P., 1991. Multivariable environmental inferences based on diatom assemblages from Sudbury (Canada) lakes. Freshwater Biology, **26**: 251-266.

Gauch, H.G., 1982. Multivariate Analysis in Community Ecology. Cambridge University Press, Cambridge: 298 pp.

Glew, J.R., 1991. Miniature gravity corer for recovering short sediment cores, Journal of Palaeolimnology, **5**: 285-288.

Grimm, E.C., 1991. TILIA version 1.11. TILIAGRAPH version 1.18. In: A. Gear (ed.), A Users Notebook. Illinois State Museum, Springfield, USA.

Hall, R.I. & Smol, J.P., 1992. A weighted-averaging regression and calibration model for inferring total phosphorus concentration from diatoms in British Columbia (Canada) lakes. Freshwater Biology, **27**: 417-434.

Hill M.O., 1973. Diversity and evenness: a unifying notation and its consequences. Ecology, **54**: 427-432.

Hill, M.O. & Gauch, H.J. (Jr.), 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio, **42**: 47-58.

Hinkley, D.V., 1983. Jackknife methods. Encyclopedia Stat. Sci, **4**: 280-287.

Hustedt, F., 1930-1966. Die Kieselalgen Deutschlands, Österreichs und der Schweiz mit Berücksichtigung der übrigen Länder Europas sowie der angrenzenden Meeresgebiete. Kryptogamen-Flora. 7 Vol. 1 (1927-1930), 2 (1937-1959), 3 (1961-1966). Geest and Portig, Leipzig.

Hustedt, F., 1937-1939. Systematische und ökologische Üntersuchungen über den Diatomeen-flora von Java, Bali, Sumatra. Archiv für Hydrobiologie (suppl.) **15-16**.

Hustedt, F., 1957. Die Diatomeenflora des Fluss-systems der Waser im Gebiet der Hansestadt Bremen. Ab. Naturw. Ver. Bremen, **34**: 181-440.

Hustedt, F., 1959. Dr. L. Rabenhordt's Kryptogamen-Flora von Deutschland die Kieselalgen, Teil 2, von Hustedt. (2nd edition; 1977 reprint). Otto Koeltz, Koenigstein.

Hutchinson, G.E., 1967. A Treatise on Limnology. Vol II. Wiley, New York, 1015 pp.

Jongman, R.H., ter Braak, C.J.F. & van Tongeren, O.F.R. (eds.), 1987. Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen: 299 pp.

Juggins, S., 1991. TRAN user manual. Unpublished.

Juggins, S. & ter Braak, C.J.F. 1992. CALIBRATE version 0.15, unpublished computer program.

Krammer, K. & Lange-Bertalot, H., 1986. Bacillariophyceae. 1 Teil: Naviculaceae. In: H. Ettl, J. Gerloff, H. Heynig & D. Mollenhauer (eds.), Süsswasserflora von Mitteleuropa. Gustav Fischer Verlag, Stuttgart: 876 pp.

Krammer, K. & Lange-Bertalot, H. 1988. Bacillariophyceae. 2 Teil: Bacillariaceae, Epithemiaceae, Surirellaceae. In: H. Ettl, J. Gerloff, H. Heynig & D. Mollenhauer (eds.), Süsswasserflora von Mitteleuropa. Gustav Fischer Verlag, Stuttgart: 596 pp.

Krammer, K. & Lange-Bertalot, H., 1991a. Bacillariophyceae. 3 Teil: Centrales, Fragilariaceae, Eunotiaceae. In: H. Ettl, J. Gerloff, H. Heynig & D. Mollenhauer (eds.), Süsswasserflora von Mitteleuropa. Gustav Fischer Verlag, Stuttgart: 576 pp.

Krammer, K. & Lange-Bertalot, H., 1991b. Bacillariophyceae. 4 Teil: Achnanthaceae Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema. In: H. Ettl, J. Gerloff, H. Heynig & D. Mollenhauer (eds.), Süsswasserflora von Mitteleuropa. Gustav Fischer Verlag, Stuttgart: 437 pp.

Mackereth, F.J.H., 1966. Some chemical observations on post-glacial sediments. Phil. Trans. R. Soc. B, **250**: 165-213.

MacDonald, G.M., Edwards, T.W.D., Moser, K.A., Pienitrz, R. & Smol, J.P., 1993. Rapid response of treeline vegetation and lakes to past climate warming. Nature, **361**: 243-246.

Manly, J.W., 1991. The design and analysis of research studies. Cambridge University Press, Cambridge, 353 pp.

PAGES workshop report, 1993. Research Protocols for PALE. Palaeoclimates of Arctic Lakes and Estuaries, series 94-1, issued by PALE steering committee.

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

Payette, S. & Lavoie, C., 1994. The Arctic tree line as a record of past and recent climatic changes. Environ. Rev., **2**: 78-89.

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

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

Psenner, R. & Schmidt, R., 1992. Climate-driven pH control of remote alpine lakes and effects of acid deposition. Nature, **356**: 781-783.

Rhee, G.Y. & Gothman, 1981. The effect of environmental factors on phytoplankton growth: temperature and the interactions of temperature with nutrient limitation. Limnol. Oceanogr., **26**: 635-648.

Rond, M.C., Greenberg, A.E. & Taras, M.I. (eds). 1975. Standard Methods for Examination for Water and Wastewater. Fourteenth edition, USA, 1195 pp.

Round, F.E., 1981. The Ecology of the Algae. Cambridge University Press, Cambridge: 653 pp.

Schindler, D.W., Curtis, P.J., Parker, B.R. & Staiton, M.P., 1996. Consequences of climate warming and lake acidification for UV-B penetration in North American boreal lakes. Nature, **379**: 705-707.

Seppa, H., 1996. Post-glacial dynamics of vegetation and tree-lines in the far north of Fennoscandia. Fennia, **174**: 1-77.

Simonsen, A., 1987. Atlas and Catalogue of the Diatom Types of Friedrich Hustedt Vols. I-III. Cramer, Berlin.

Smol, J.P, 1988, Palaeoclimate proxy data from freshwater diatoms. Vern. Int. Ver. Limnol., **23**: 837-844.

Smol, J.P., Walker, I.R. & Leavitt, P.R., 1991. Palaeolimnology and hindcasting climatic trends. Vern. Int. Ver. Limnol. **24**: 1240-1246.

Stevenson, A.C., *et al.*, 1991. The Surface Waters Acidification Project Palaeolimnology Programme: Modern Diatom/Lake-water Chemistry Data-set. ENSIS Publishing, London, England: 86 pp.

ter Braak, C.J.F. 1983. Principal component biplots and alpha and beta diversity. Ecology, **64**: 454-462.

ter Braak, C.J.F., 1987a. The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio, **69**: 69-77.

ter Braak, C. J. F., 1987b. Calibration. In: R.H.G. Jongman, C.J.F. ter Braak & O.F.R. van Tongeren (eds.), Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen: pp 78-90.

ter Braak, C.J.F., 1987c. Ordination. In: R.H.G. Jongman, C.J.F. ter Braak & O.F.R. van Tongeren (eds.), Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen: pp 91-173.

ter Braak, C.J.F., 1988a. CANOCO - A FORTRAN program for canonical community ordination by [partial] [detrended][canonical] correspondence analysis, principal component analysis and redundancy analysis (version 2.1.). Technical report LWA-88-02. GLW, Wageningen.

ter Braak, C.J.F., 1988b, Partial canonical correspondence analysis. In: H.H. Lock (ed.), Classification and related methods of data analysis, North Holland, Amsterdam: pp 551-558.

ter Braak, C.J.F., 1990. Update notes: CANOCO version 3.10. Agricultural Mathematics Group, Wageningen: 35 pp.

ter Braak, C.J.F., 1995. Canonical community ordination. Part 1: Basic theory and linear methods. Ecoscience, **1**: 120-140.

ter Braak, C.J.F. & Barendregt, L.G., 1986. Weighted averaging of species indicator values: its efficiency in environmental calibration. Math. Biosci., **78**: 57-72.

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

ter Braak, C.J.F. & Looman, C.W.N., 1986. Weighted averaging, logistic regression and the Gaussian response model. Vegetatio, **65**: 3-11.

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

ter Braak, C.J.F. & van Dam, H., 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. Hydrobiologia, **178**: 209-223.

Tilman, D. & Kiesling, R.L., 1984. Freshwater algal ecology: Taxonomic trades-off in the temperature dependence of nutrient competitive abilities. In: M.J. Klug & C.A. Reddy, (eds), Proceedings of the 3rd International Symposium on Microbial Ecology, Michigan State University, pp. 314-319.

Vyverman, W., 1992. Altitudinal distribution of non-cosmopolitan desmids and diatoms in Papua New Guinea. Brit. Phycol. J., **27**: 49-63.

Vyverman, W. & 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, Journal of Paleolimnology, **13**: 65-77.

Wright, H.E. Jr., 1984. Sensitivity and response time of natural systems to climatic change in the late Quaternary, Quatern. Sci. Rev., **3**: 91-131.

Appendix 1 Study area and site descriptions

Lake	Location	Altitude (m)	Max. depth (m)	Lake area (km^2)	Geology	Vegetation zone
KOLA1	33°48'; 67°33'	145	5	0.04	granite	mixed taiga
KOLA2	33°28'; 67°42'	450	8	0.04	ultra-alkaline	alpine birch forest/tundra
KOLA3	35°37'; 68°51'	110	1.5	0.04	granite	birch forest/tundra
KOLA4	35°24'; 68°48'	150	7	0.04	granite	birch forest/tundra
KOLA5	32°57'; 69°14'	101	19.2	0.04	granite	birch forest/tundra
KOLA6	32°53'; 69°03'	160	2.6	0.04	granite	birch forest/tundra
KOLA7	32°37'; 69°11'	240	7	0.04	granite	birch forest/tundra
KOLA8	34°59'; 69°11'	170	5.5	0.3	granite	tundra
KOLA9	34°53'; 68°53'	230	1.5	0.08	granite	tundra
KOLA10	34°53'; 68°57'	260	4	0.04	granite	tundra
KOLA11	34°58'; 69°3'	110	12	0.28	granite	tundra
KOLA12	68°55'; 34°0'	218.7	2.7	0.09	granite	birch forest/tundra
KOLA13	31°07'; 68°44'	104	3.5	0.04	gneiss, schist	pine taiga
KOLA14	30°34'; 68°43'	120	6.2	0.04	gneiss, schist	pine taiga
KOLA15	28°41'; 68°38'	150	4.2	0.11	gneiss, schist	pine/birch taiga
KOLA16	28°50'; 68°31'	105	4.1	0.04	gneiss, schist	pine taiga
KOLA17	29°17'; 68°32'	110	5	0.04	granite, sand	pine taiga
KOLA18	30°41'; 68°43'	111.5	5	0.04	gneiss, schist	spruce taiga
KOLA19	33°29'; 67°34'	150	3	0.08	granite	spruce taiga
KOLA20	33°12'; 68°42'	145	4	0.05	granite	birch forest/tundra
KOLA21	33°24'; 68°46'	151	4	0.08	granite	birch forest/tundra
KOLA22	35°53'; 69°2'	117	6.5	0.04	granite, sand	tundra
KOLA23	36°4'; 69°4'	79	6.7	0.04	granite	tundra
KOLA24	35°57'; 69°3'	125	18.5	0.04	granite	tundra
KOLA25	32°27'; 67°57'	506	16	0.04	granite	alpine/tundra

Appendix 2 Water chemistry of the modern training set (mean values)

SITE	pH	COND ($\mu S\ cm^{-1}$)	COL (°Pt)	NH_4 ($\mu g\ l^{-1}$)	N_{tot} ($\mu g\ l^{-1}$)	TOC ($mg\ l^{-1}$)	P_{tot} ($\mu g\ l^{-1}$)	$Si0_2$ ($mg\ l^{-1}$)	COD ($mg\ l^{-1}$)	ALK $\mu eq\ l^{-1}$)
KOLA1	7.4	41	5	15	143	2.32	7	0.34	1.4	283
KOLA2	7.0	47	5	15	388	2.5	10	2.9	11	210
KOLA3	5.0	35	140	24	470	10.4	17	0.75	11.6	-9
KOLA4	6.3	29	149	8	275	10.1	7	0.16	11.2	56
KOLA5	6.7	43	5	1	110	2.8	3	1	1.67	61
KOLA6	6.4	88	95	3	284	9.2	11	0.24	10	62
KOLA7	6.2	30	65	3	187	7.1	5	1.28	7.3	36
KOLA8	6.4	35	16	9	94	3.4	6	0.65	2.45	49
KOLA9	6.4	28	46	28	215	5.7	8	1.12	5.4	59
KOLA10	6.3	39	27	3	133	4.2	6	0.53	3.44	27
KOLA11	6.4	33	38	4.6	208	5.2	10	0.32	4.8	37
KOLA12	6.7	34	31	2	240	5.3	10	1.13	4.9	74
KOLA13	6.9	34	57	7	248	7.1	4	0.93	7.2	200
KOLA14	6.3	19	110	4	216	9.6	12	1.74	10.6	58
KOLA15	6.9	23	88	16	305	9.6	5	0.53	10.5	127
KOLA16	6.9	16	3	8	200	3.1	8	0.89	2.08	96
KOLA17	5.1	8	7	6	256	5.1	6	0.04	4.7	-4
KOLA18	6.7	29	120	7	249	10.1	5	2.37	11.2	131
KOLA19	5.1	23	73	3	224	7.4	7	0.17	7.6	-2
KOLA20	6.3	32	240	21	367	20.2	9	2.91	24.4	79
KOLA21	6.3	30	157	16	229	10.7	5	0.97	12	54
KOLA22	5.8	30.5	26	9	126	3.9	7	0.43	3.05	10
KOLA23	6.3	42	40	14	134	4.9	4	0.66	4.44	33
KOLA24	6.6	41	38	6	116	4.4	2	0.92	3.72	53
KOLA25	6.4	9	4	7	234	0.73	12	0.7	1.63	24.3

Appendix 2 Cont.

SITE	K ($mg\ l^{-1}$)	Na ($mg\ l^{-1}$)	Ca ($mg\ l^{-1}$)	Mg ($mg\ l^{-1}$)	SO_4 ($mg\ l^{-1}$)	Cl ($mg\ l^{-1}$)	Fe ($\mu g\ l^{-1}$)
KOLA1	0.99	3.32	4.18	0.51	5.0	0.95	1.8
KOLA2	1.17	5.50	0.50	1.00	1.40	0.80	2
KOLA3	0.27	4.53	0.39	0.71	2.55	7.63	110
KOLA4	0.23	3.58	1.22	0.63	1.44	5.51	320
KOLA5	0.41	4.88	1.52	0.91	4.71	8.44	75
KOLA6	0.83	11.30	2.48	1.23	3.14	21.90	118
KOLA7	0.34	3.09	1.04	0.66	3.30	4.82	74
KOLA8	0.24	4.45	1.09	0.71	2.23	7.68	35
KOLA9	0.24	3.19	1.09	0.60	2.71	4.96	20
KOLA10	0.28	5.31	0.67	0.72	2.20	9.41	24
KOLA11	0.30	4.37	0.81	0.63	2.20	7.24	40
KOLA12	0.26	3.63	1.22	0.89	1.57	6.33	19
KOLA13	0.82	1.63	2.40	1.40	2.82	1.51	28
KOLA14	0.42	1.78	2.39	0.94	2.04	1.94	135
KOLA15	0.35	1.28	1.87	0.91	1.51	1.12	80
KOLA16	0.31	0.91	1.13	0.48	1.13	0.86	23
KOLA17	0.13	0.44	0.30	0.14	1.06	0.65	21
KOLA18	0.41	1.73	2.87	0.80	2.72	1.82	190
KOLA19	0.25	1.35	1.74	0.26	5.00	1.67	41
KOLA20	0.48	2.93	1.92	1.36	3.30	3.71	195
KOLA21	0.25	3.63	1.18	0.70	1.26	5.50	84
KOLA22	0.23	3.86	0.37	0.56	1.90	6.93	38
KOLA23	0.31	5.64	0.74	0.82	2.53	9.43	52
KOLA24	0.30	5.42	1.00	0.82	2.53	8.84	35
KOLA25	0.04	0.50	0.90	0.14	1.80	0.62	49

Appendix 3 Estimated WA pH optima of the 99 diatom taxa, their number of occurrence, effective number of occurrence (N2) as well as their taxon code used in Figures

No	Taxon code	Taxon name	# Occ	N2	Opt.	Tol
1.	ACPSEU	*Achnanthes pseudoswazi* Carter	5	3.47	6.55	0.32
2.	ACMINU	*Achnanthes minutissima* var. *minutissima* Kutz.	20	9.70	6.63	0.50
3.	ACNODO	*Achnanthes nodosa* Cleve-Euler	14	7.88	6.59	0.57
4.	ACSUCH	*Achnanthes suchlandtii* Hust.	4	1.49	7.17	1.13
5.	ACPUSI	*Achnanthes pusilla* var. *pusilla* Grun.	14	8.18	6.33	0.70
6.	ACDIDY	*Achnanthes didyma* var. *didyma* Hust.	5	3.00	6.88	0.82
7.	ACDETH	*Achnanthes detha.*	11	3.15	6.63	0.74
8.	ACLEVA	*Achnanthes levanderi* Hust.	5	3.20	6.86	0.63
9.	ACALTA	*Achnanthes altaica* (Poretzky) Cleve-Euler	19	12.16	6.32	0.53
10.	ACCURT	*Achnanthes curtissima* Carter	12	7.41	6.76	0.44
11.	ACKUEL	*Achnanthes kuelbsii* Lange-Bertalot	14	8.05	6.21	0.61
12.	ACLVUL	*Achnanthes lacus-vulcani* Lange-Bertalot & Kramer	8	4.84	6.52	1.02
13.	ACMARG	*Achnanthes marginulata* Grun in Clev.& Grun	18	7.93	6.51	0.32
14.	ACSCOT	*Achnanthes scotica* Jones & Flower	12	8.60	6.52	0.35
15.	ACKRIE	*Achnanthes kriegeri* Krasske	8	2.69	6.80	0.36
16.	ACFLEX	*Achnanthes flexella* (Kutz.) Brun	7	4.97	6.25	0.73
17.	ACHELV	*Achnanthes austriaca* var. *helvetica* Hust.	7	5.22	6.32	0.19
18.	ACSPEC	*Achnanthes sp.*	4	1.98	6.37	0.87
19.	AUALPE	*Aulacoseira lirata* var. *alpigena* (Grun) Haworth	18	8.06	6.48	0.51
20.	AULACU	*Aulacoseira lirata* var. *lacustris* (Grun. in Van Heurck) Ross in Hartley	8	1.38	6.29	0.33
21.	AULILA	*Aulacoseira lirata* var. *lirata* (Ehrenb.) Ross in Hartley	4	2.67	6.42	0.23
22.	AUFLOR	*Aulacoseira perglabra* var. *floriniae*	11	6.88	6.27	0.66
23.	AUPERG	*Aulacoseira perglabra* var. *perglabra*	16	7.43	6.19	0.33
24.	AUVALI	*Aulacoseira italica* var. *valida* (Grun. in Van Heurck) Simonsen	7	3.13	6.24	0.44
25.	AUSUBA	*Aulacoseira italica* var. *subarctica* (O.Mull) Simonsen	9	3.72	6.42	0.33
26.	AUNIVA	*Aulacoseira distans* var. *nivalis* (Nygaard) Ross	15	3.97	6.08	0.38
27.	AUTENE	*Aulacoseira distans* var. *tenella* (Nygaard) Ross	8	5.15	5.81	0.70
28.	AUNYGA	*Aulacoseira nygaardii* Camburn	4	2.73	5.98	0.85
29.	APRUTI	*Amphipleura rutilans* (Trentepohl ex Roth) Cleve	5	4.17	6.62	0.34
30.	BRVITR	*Brachysira vitrea* (Grun.) Ross in Hartley	23	13.95	6.49	0.43
31.	BRBREB	*Brachysira brebissonii* var. *brebissonii* Ross	24	13.23	6.31	0.47
32.	BRZELL	*Brachysira zellensis* (Grun) Round & Mann	7	2.46	6.82	0.25
33.	CMPERP	*Cymbella perpusilla* Cleve	16	10.64	6.21	0.75
34.	CMCESA	*Cymbella cesatii* var. *cesatii* (Rabenh.) Grun in Schmidt	14	5.29	6.80	0.24
35.	CMMINU	*Cymbella minuta* var. *minuta* Hilse ex Rabenh	10	3.38	7.02	0.42
36.	CMLUNA	*Cymbella lunata* W.Sm. in Grev.	22	13.77	6.45	0.39
37.	CMMICR	*Cymbella microcephala* var. *microcephala* Grun in Van Heurck	9	5.29	6.64	0.31
38.	CMHERB	*Cymbella hebridica* (Grun ex Clev.) Cleve	16	10.26	6.14	0.51
39.	CMGAEM	*Cymbella gaeumannii* Meister	13	7.93	6.46	0.22
40.	CMDESC	*Cymbella descripta* (Hust) Kram. & Lange-Bert.	7	4.18	6.72	0.30
41.	CMNAVI	*Cymbella navicoliformis* Auersw. ex Heib	9	5.56	6.20	0.75
42.	CYPSEU	*Cyclotella pseudostelligera* Hust	4	1.67	7.25	0.52
43	CYROSS	*Cyclotella rossii* Hakansson	7	3.09	6.61	0.18

Appendix 3 Cont.

No	Taxon code	Taxon name	# Occ	N2	Opt.	Tol
44.	EUCURV	*Eunotia curvata* var. *curvata* (Kutz) Lagerst.	8	5.22	6.18	0.73
45.	EUSPEC	*Eunotia sp.*	16	5.25	6.04	0.74
46.	EURHOM	*Eunotia rhomboidea* Hust.	9	12.0	6.34	0.41
47.	EUSERR	*Eunotia serra* var. *serra* Erhenb.	10	5.66	6.18	0.63
48.	EUPRAE	*Eunotia praerupta* var. *praerupta* Ehrenb.	10	2.17	6.73	0.47
49.	EUVANH	*Eunotia vanheurckii* var. *vanheurckii* Patr.	21	7.47	6.31	0.30
50.	FRVENT	*Fragilaria construens* var. *venter* (Ehrenb) Grun. in Van Heurck	10	5.70	6.61	0.71
51.	FRVIRE	*Fragilaria virescens* var. *exigua* Grun in Van Heurck		11.53	6.29	0.55
52.	FRELLI	*Fragilaria elliptica* Schum.	7	4.11	6.77	0.70
53.	FRPSEU	*Fragilaria pseudoconstruens* Marciniak	7	3.79	6.52	0.64
54.	FRCBRE	*Fragilaria [cf. brevistriata]* Botungen	9	2.46	6.96	0.84
55.	FRCNRI	*Fragilaria constricta* var. *constricta* Ehrenb	9	7.36	6.22	0.39
56.	FRPINN	*Fragilaria pinnata* var. *pinnata* Ehrenb.	9	3.90	5.82	0.85
57.	FRBICA	*Fragilaria bicapitata* A.Mayer	5	3.89	6.41	0.26
58.	FUSAXO	*Frustulia rhomboides* var. *saxonica* (Rabenh) De Toni	23	11.24	6.01	0.65
59.	FURHOM	*Frustulia rhomboides* var. *rhomboides* (Ehrenb) De Toni	18	8.25	6.02	0.56
60.	FUVIRI	*Frustulia rhomboides* var. *viridula* (Breb ex Kutz) Cleve	14	10.4	6.18	0.54
61.	GOPARV	*Gomphonema parvulum* var. *parvulum* Kutz.	10	5.71	6.59	0.53
62.	MEAREN	*Melosira arentii* (Kolbe) Nagumo & Kobayasi	6	3.57	6.45	0.67
63.	NARADI	*Navicula radiosa* var. *radiosa* Kutz.	9	5.50	6.74	0.56
64.	NAPSEU	*Navicula pseudoscutiformis* Hust.	8	2.12	6.76	0.97
65.	NAPUPU	*Navicula pupula* var. *pupula* Kutz.	9	2.15	6.38	0.66
66.	NABEGE	*Navicula begeri* Krasske	13	4.05	6.05	0.82
67.	NADIGI	*Navicula digitulus* Hust.	6	6.27	6.78	0.44
68.	NAHOEF	*Navicula hoefleri* Sensu Ross et Sims	19	15.01	5.81	0.86
69.	NAMEDI	*Navicula mediocris* Krasske	19	12.13	6.31	0.50
70.	NAJAAG	*Navicula jaagii* Meister	5	2.71	6.79	0.52
71.	NALEPT	*Navicula leptostriata* Jorgensen	20	13.65	6.33	0.47
72.	NASUBT	*Navicula subtilissima* Cleve	17	6.50	5.83	0.79
73.	NAIMPE	*Navicula impexa* Hust.	5	3.57	6.49	0.32
74.	NASUBM	*Navicula subminuscula* Manguin	3	2.57	6.11	0.30
75.	NAVITI	*Navicula vitiosa* Schimanski	5	1.63	6.76	0.33
76.	NACFIM	*Navicula [cf. impexa]* Scoat Tarn (EYH-SWAP)	3	2.44	5.69	0.78
77.	NACOCC	*Navicula cocconeiformis* var. *cocconeiformis* Greg. ex Greville	5	1.96	6.41	0.17
78.	NASOEH	*Navicula soehrensis* var. *soehrensis* Krasske	9	5.20	6.33	0.22
79.	NAPVEN	*Navicula pseudoventralis* Hust.	3	2.59	6.47	0.25
80.	NASERA	*Navicula seminulum* var. *radiosa*	3	1.76	6.58	0.26
81.	NATENU	*Navicula tenuicephala* Hust.	5	3.09	5.67	0.82
82.	NASEMI	*Navicula seminulum* Grun	4	2.13	6.51	0.85
83.	NAMINU	*Navicula minuscula* Grun in Van Heurck	4	1.96	6.71	0.39
84.	NIFONT	*Nitzchia fonticola* Grun in Van Heurck	19	9.15	6.56	0.41
85.	NIPERM	*Nitzchia perminuta* (Grun in Van Heurck) M.Perag	22	13.81	6.55	0.62
86.	NIPALE	*Nitzchia palea* var. *palea* (Kutz.) W.Sm.	3	2.32	6.71	0.29
87.	NIREST	*Nitzchia recta* Hantzch ex Rabenh.	12	8.78	6.25	0.61
88.	NICGRA	*Nitzchia [cf. gracilis]* SWAP Sweden	6	3.73	6.58	0.48