Defining restoration targets for acidified upland lakes using diatom and cladoceran sub-fossil remains and the modern analogue technique

Thesis submitted for the degree of Doctor of Philosophy
University of London

Gavin Leslie Simpson

Department of Geography
University College London
December 2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Evaluating recovery of acidified lakes towards a target based on their pre-acidification status is a difficult task as there are very few biological records for these remote upland systems. The analogue matching approach can be used to identify modern analogues for the pre-acidification status of lakes using the sub-fossil remains of diatoms and Cladocera, which can be used as recovery targets.

An 83-lake modern training set was created, with full diatom, cladoceran, hydrochemistry and catchment data for each lake. Fossil samples from 10 lakes of the UK Acid Waters Monitoring Network (UKAWMN) were chosen to represent pre-acidification conditions in those lakes. The closest analogues in the modern training set were identified for each fossil sample by means of the squared chord distance measure.

The distributions of the Cladocera in the training set were analysed using a range of multivariate statistical technique. Individual cladoceran species response curves for selected environmental determinands were calculated. A 163-lake diatom training set has been compiled from existing data holdings and the patterns in the data were explored using direct ordination methods. A comparison with the diatom training set from the Surface Waters Acidification Project is also made.

Close modern analogues were identified for 8 UKAWMN lakes. The majority of these modern analogues are located in North and Northwest Scotland; areas of low sulphur and nitrogen deposition. Comparison of the hydrochemical characteristics of the UKAWMN lakes to those of the modern analogues showed that the modern analogues had higher lake water pH and alkalinity levels and lower aluminium concentrations. Ionic strength and calcium concentrations in the analogue lakes were similar to observed values in the UKAWMN lakes.

These results indicate that the analogue matching approach using diatom and cladoceran remains is a simple, robust and reliable method of identifying modern analogues for acidified lakes in upland areas of the UK.

Acknowledgments

This research was supported by ENSIS Ltd. as part of the Critical Loads of Acidity and Metals (CLAM) project funded by the Department of Environment, Food and Rural Affairs (DEFRA contract EPG/1/3/117). Funding to attend Cladocera taxonomic workshops in Helsinki, Finland was provided by the UCL Graduate School and the Department of Geography's Mead travel fund.

I would like to thank my supervisors Tim Allott (Manchester) and Rick Battarbee for all their help and advice over the past four years. In particular I would like to thank Rick for his enthusiasm for my research topic and for pointing out that zooplankton research was not sufficiently advanced to tackle cladoceran *morality* yet. I await the day!

Numerous other people have assisted my work in many ways:

- The CLAM Boys (Chris Curtis, Martin Kernan, Don Monteith, Roger Flower, Neil Rose, Simon Patrick and Mike Hughes) deserve particular mention for their useful contributions to my work, stimulating discussion, payment of expense claims and general banter. Special thanks to Chris for the friendly competition and underhand tactics as we both struggled to write up! What do I owe you?

- Janet Hope, Tula Maxted and Ian Patmore for their support in the laboratory and for devising ever more devilish ways of extracting cladoceran remains from dried, clay rich sediments.

- The ECRC; I have had the great pleasure of working in such a friendly and thought provoking environment for the past four years. I am indebted to everyone in the ECRC, past and present, that I was fortunate enough to meet and work with.

- John Birks (Bergen & UCL), for throwing me in at the 'statistical deep end' 6 months into my studies, for your advice and suggestions on data analysis, and for providing a specially compiled version of ANALOG.

For their constant support in so many ways throughout my life and in particular over the past seven years whilst I have been a student at UCL I would like to thank my parents and brother, Matthew.

Finally, Marianne, without your love and encouragement over the years and your job as sounding board this thesis would not have been possible. From the bottom of my heart, I thank you.

Table of contents

Detailed table of contents

List of Figures

List of Tables

Chapter 1: Lake acidification, recovery, and the modern analogue approach

1.1 Acid deposition, surface water acidification and palaeolimnology

The acidification of surface waters via atmospheric contamination by strong acid-forming compounds has been a major environmental issue for the last 30 years or so. Angus Smith (1852 cited in Battarbee *et al.* 1988a) first coined the term 'acid rain' when he used it to describe the effect of coal combustion on air and precipitation chemistry at various industrial cities in the UK. The effect this might have on lake ecosystems was first discussed by Gorham (1958). Since then numerous studies have been published which have investigated the problems associated with acid deposition and have confirmed that lake acidification has taken place in large areas across Europe and North America (see for example Battarbee and Charles 1986). Studies have shown extensive acidification of many lakes within these regions. For example, the Palaeoecological Investigation of Recent Lake Acidification (PIRLA) (Charles and Whitehead 1986) was a large-scale project that attempted to assess the extent and cause of acidification in North America. In the Adirondack Mountains National Park, New York, all 11 clear-water lakes from the region sampled as part of PIRLA, where contemporary pH was below 5.5, were found to have been acidified as a result of the deposition of strong acids on to their catchments.

The counterpart in Europe to PIRLA was the Palaeolimnology Programme of the Surface Waters Acidification Project (SWAP) (Battarbee and Renberg 1990). The Palaeolimnology Programme was planned in 1984 following a period of extensive research on the recent acidification of surface waters in Europe and North America. The Palaeolimnology Programme involved forty scientists from the UK, Sweden and Norway and forged close links with the PIRLA project studying surface water acidification in North America.

One of the main features of the SWAP Palaeolimnology Programme was a series of projects designed to test a variety of proposed mechanisms for the cause of surface water acidification in Northern Europe and North America, such as Rosenqvist's (1978) land use change hypothesis, the magnitude of long-term acidification processes (Pennington 1984), and the role afforestation played in the acidification (Stoner, Gee, and Wade 1984). The second major part of the programme was the integration of detailed palaeolimnological studies in Norway, Sweden and the UK. A wide range of biological proxies and pollution indicators were examined at sites in areas of high and low deposition in an attempt to characterise the patterns, trends, timings and the impacts of acid deposition on surface waters in Northern Europe (for an overview see Battarbee and Renberg 1990; Battarbee 1994).

The main findings of the SWAP Palaeolimnology Programme were that there was a lack of evidence to support the alternative hypotheses as the cause of surface water acidification. The programme showed that there was overwhelming palaeolimnological evidence to support the proposition that acid deposition was the main causal factor in the recent acidification of surface waters in North Europe.

In the UK, Battarbee *et al.* (1988a) identified a range of lakes impacted by atmospheric contamination resulting in increased acidity levels. Acidification of UK surface waters has been identified in central and northern Wales at sites on Lower Palaeozoic Sedimentary and Metamorphic rocks. Acidification has also been identified in Cumbria, on Borrowdale volcanic strata, and in Scotland, in Galloway, the Trossachs and the Grampians. There are also numerous highly acidic ponds and lakes in the English Pennines. Since this work, further studies have shown that lakes have slightly acidified in areas previously thought to be unaffected by atmospheric deposition, such as remote sites in northwest Scotland (Allott, Golding, and Harriman 1995).

Palaeolimnological studies (e.g. Charles and Whitehead 1986; Battarbee and Renberg 1990) have proved conclusively that acid deposition has resulted in the widespread acidification of surface waters. The timing of acidification in lakes has always occurred after the onset of the Industrial Revolution and the record of carbonaceous particle deposition in lake sediments clearly demonstrates atmospheric contamination from industrial sources in parallel to acidification (Rose *et al.* 1995; Rose 1996).

1.1.1 International Emissions Reductions

The acceptance of the cause-effect relationship between acid deposition and acidification has led to international efforts to reduce acid emissions. In 1979, the UNECE Convention on Long-Range Transboundary Air Pollution (LRTAP) was adopted to implement measures to reduce the levels of sulphur dioxide emissions from industrial sources. Currently there are 40 signatory countries to LRTAP, which has identified principles for international co-operation and for the abatement of the emissions of pollutants.

Under the LRTAP convention, a number of protocols have been developed that commit member states to certain abatement actions. The early protocols limited emissions of sulphur dioxide (the 1985 Helsinki Protocol) and oxides of nitrogen (the 1988 Sofia Protocol). All countries were required to cap their emissions or reduce them in relation to a given reference year (see Table 1). In 1994, the Second Protocol on the Further Reduction of Sulphur Emissions was adopted. This, the Oslo Protocol, introduced a new method of setting reduction targets using critical loads, an effects approach to deposition (Bull 1995) based on critical loads maps that represent the sensitivity of ecosystems to given levels of acid deposition. Using modelled deposition across Europe for given abatement strategies, critical load exceedance maps are used to indicate the effect on ecosystems for that level of abatement. Setting abatement targets for international protocols now involves the benefits to ecosystems that a given reduction will involve, and through Integrated Assessment Modelling cost effective abatement strategies have been developed.

The latest UNECE LRTAP protocol was adopted at a meeting in Gothenburg in 1999. This, the Gothenburg protocol or the "*multi-pollutant multi-effect protocol*" is unique in that it calls for the reduction of four pollutants (sulphur dioxide, nitrogen dioxide, ammonia and ground level ozone) to reduce the effects of three environmental problems (acidification, eutrophication, and tropospheric ozone). The protocol aims to decrease critical load exceedance in a cost effective manner by 2010 based on a gap closure methodology. Under the Gothenburg protocol the UK is now committed to capping annual emissions of sulphur dioxide at 625 kilo tonnes (ktonnes), emissions of nitrogen oxides at 1181 ktonnes and emissions of ammonia at 297 ktonnes (NEGTAP 2001).

Table 1: Current protocols developed as part of the UNECE Convention on Long-Range Transboundary Air Pollution and the European Union pollution emission control protocols to curb the emissions of sulphur dioxide and oxides of nitrogen implicated in the acidification of surface waters across Europe and North America.

Protocol	Year Adopted	Year in Force	Requirements of the Protocol
UNECE CONVENTION ON LONG-RANGE TRANSBOUNDARY AIR POLLUTION PROTOCOLS			
EMEP	1984	1988	The Protocol on Long-Term Financing of the Co-operative Programme for Monitoring and Evaluation of Long-Range Transmission of Air Pollutants in Europe. Collates information on deposition and emission inventories supplied by member states, and develops transport models for pollutants.
Helsinki	1985	1987	The Protocol on the reduction of sulphur emissions or their transboundary fluxes by at least 30%. Committed parties to a 30% cut in SO2 emissions by 1993 based on 1980 levels.
Sofia	1988	1991	The Protocol concerning the control of emissions of nitrogen oxides or their transboundary fluxes. Commits member states to return NOx emissions to their 1987 levels by 1994
Oslo	1994	1998	The Second Protocol on the further reduction of sulphur emissions. requires different % reductions from member states based upon an effects based concept know as the critical loads approach. The UK is required to reduce sulphur emissions by 80% against 1980 levels, by 2010.
Gothenburg	1999	-	Acidification, Eutrophication and Ground-level Ozone (Gothenburg Protocol). Calls for reductions by 2010 over 1990-levels of sulphur dioxide (75%), Nitrogen Oxides (50%), non Methane VOCs (58%) and Ammonia (12%).
EUROPEAN UNION POLLUTION EMISSION CONTROL PROTOCOLS			
5EAP	1993	1993	5th Environmental Action Programme called for reductions in sulphur dioxide (35% of 1985 levels by 2000), nitrogen oxides (30% of 1990 levels by 2000), non-methane VOCs (30% of 1990 levels by 1999).
NECD	1999	-	National Emissions Ceilings Directive (NECD) calls for reductions in sulphur dioxide (77%), nitrogen oxides (55%), non-methane VOCs (54%) and ammonia (14%) over 1990 levels by 2010.

European Union emission control protocols have also been developed that bind member states to reductions in the emission of certain air pollutants. The National Emissions Ceilings Directive (NECD) is one such protocol and calls for reductions in sulphur dioxide, nitrogen dioxide, ammonia and ground level ozone to 1990 levels by 2010 and is similar in many respects to the Gothenburg protocol.

In the light of these international reductions in air pollutants, the emphasis is now shifting towards the recovery of acidified surface waters. There is now a need to investigate the process of recovery in acidified surface waters to evaluate and model the response of surface waters to reduced acid deposition. Given that the protocols arranged under the LRTAP convention are having success in reducing the level of acid deposition to

catchments across Europe and North America the role of recovery is now central to the acid deposition debate.

As part of the UK Government's commitments to reduce emissions of acid forming compounds under the various UNECE protocols and EU directives the UK Acid Waters Monitoring Network (UKAWMN) has been established specifically to identify chemical and biological responses to these reduced emissions. The network was established in 1988 and comprises 22 lake and stream sites across the UK. Further information on the UKAWMN can be found in Section 5.4.1 below.

1.2 The Biological Effects of Acidification

The changes in the acidity of acidified lakes have had important consequences for their biology. Although effects on fisheries and salmonids in particular have received most attention algae, plants, invertebrates and other animals have also been affected. Biota can be affected directly by episodic acid events and long term, sustained chronic acidification, and indirectly by changes in food webs leading to changes in grazing and predation pressure (Muniz 1991).

1.2.1 Phytoplankton and Periphyton

Primary productivity in lakes can be drastically altered by acidification; even relatively small water chemistry changes can produce large changes in phytoplankton communities (e.g. Round 1990; Vinebrooke *et al.* 2002).

Acidified lakes have fewer species and contain different dominant species than non-impacted communities. The numbers of phytoplanktonic species typically fall from 30-80 in circumneutral lakes, to 10-20 species in acidic lakes (Muniz 1991). The reasons for this shift in dominance are poorly understood. Havens & DeCosta (1987), suggest that reason for the absence of acid-tolerant species at higher pH levels is that acid-tolerant species are less competitive in less acidic conditions. The ability of acid tolerant species to dominate the phytoplankton communities in strongly acidified lakes may also be attributable to the possible protection afforded them by their morphology (e.g. chrysophytes). Jansson *et al.*

(1986) and Smith (1990) suggest that some acid tolerant species are able to withstand exposure to strongly acid waters because they have an acid phosphatase metabolism that hydrolyses organic phosphorus during periods when soluble reactive phosphate concentrations are low (Smith 1990). Alkaline phosphatase is inefficient below pH 5.5, which might explain the absence of some diatom taxa from very acidic surface waters (Smith 1990).

Synoptic surveys suggest that there is less phytoplankton biomass in acid lakes (Siegfried, Bloomfield, and Sutherland 1989), yet experimental acidification studies suggest that there are slight increases in biomass levels (Yan and Stokes 1978; Schindler *et al.* 1985). For primary productivity most studies suggest a slight increase in biomass that is possibly the result of improving light conditions extending to greater depths in acid lakes (Muniz 1991).

A feature of many non-impacted, upland lakes in the UK is a planktonic diatom flora consisting mainly of species from the genus *Cyclotella* (Round 1990). These centric diatoms can contribute a large proportion of the diatom algal population of a lake and are the dominant forms in many systems. *Cyclotella* species are missing from severely acidified lakes such as those in the Galloway region of Southern Scotland (Flower and Battarbee 1983; Flower, Battarbee, and Appleby 1987). A flora dominated by acid tolerant diatoms, *Tabellaria binalis* or *T. quadriseptata* for example, is usually found in these strongly acidified lakes. These acid tolerant diatoms live in the benthos attached to rocks, aquatic plants and associated with sediments. There is, therefore, a clear habitat change from phytoplankton in circumneutral surface waters to acid tolerant benthic communities in acidified systems. Planktonic diatoms are generally not found in acidified lakes (Round 1990). Palaeolimnological studies of acid lakes throughout Northern Europe have shown that the loss of planktonic forms is the one of the earliest biological responses to acidification (Battarbee 1984).

1.2.2 Aquatic Macrophytes

Aquatic macrophytes are a source of food for many aquatic organisms. They also provide shelter and breeding or nursery grounds. As with the observed phytoplankton community changes, acidification induced changes in the macrophyte communities could have wide-ranging effects on the biological functioning of lakes.

The evidence from North America is contradictory, however. Jackson & Charles (1988) reported decreased numbers of macrophyte species with decreasing pH and related factors. This has not been found in other studies (e.g. Yan *et al.* 1985).

Increasing acidity is linked to the large expansion of *Sphagnum* species (Grahn 1986) and, to a lesser extent, *Juncus bulbosus.* These species restrict *Lobelia dortmanna* and *Littorella uniflora,* two macrophyte species that tend to dominate the macrophyte communities in circumneutral lakes (Roelfs 1983). *Sphagnum* exchange metabolically produced hydrogen ions with nutrient and metal ions (Clymo 1984) and may, therefore, act as sinks for metals and locally acidify their microenvironments (Muniz 1991). This would reduce the area of a lake to which fish and zooplankton could migrate to avoid episodic low pH events.

1.2.3 Zooplankton

Zooplankton are directly affected by toxic conditions and indirectly by changes in the activities of primary producers at lower trophic levels and in predators from higher up the food chain. The main roles that zooplankton play in lake ecosystems are in the transfer of energy through the food chain and the recycling of nutrients (Muniz 1991). Species diversity of zooplankton communities decreases with increased acidity, especially below pH 5.5-5.0 (Locke and Sprules 1994).

Changes in sensitive zooplankton species, such as the daphnids, are often due to direct changes in the ambient water quality of acidified lake. The death of daphnids from acid stress is linked to a number of physiological stresses induced by acid toxicity; the depression of oxygen intake (Alibone and Fair 1981) and the failure of osmoregulatory processes shown in the net loss of body ions (e.g. sodium and chlorine) (Potts and Fryer 1979). Some zooplankton species are sensitive to toxic aqueous aluminium or low pH (Havens and DeCosta 1987), while others are not (Havas and Likens 1985).

Acidification is also likely to cause a decrease in predation pressure on zooplankton communities, leading to an increase in the proportion of large-bodied individuals, in acid lakes in Europe (Stenson and Eriksson 1989). In North America, smaller bodied individuals dominate zooplankton communities possibly because of more intense

predation of large bodied individuals, in preference over smaller-bodied individuals, by fish and ducks that use visual cues for hunting (Dillon, Yan, and Harvey 1984).

Experiments have shown that recovery of zooplankton communities can occur, but that they do so slowly (Hultberg and Andersson 1982). In the lakes near Sudbury, Ontario, Canada, the reversal of acidification caused by deposition of locally derived acidity pollution allowed the recolonisation of the lakes by acid-sensitive species, such as *Epischura lacustris* (Gunn and Keller 1990).

1.2.4 Benthic Macroinvertebrates

Benthic macroinvertebrates (BMIs) include molluscs (snails, clams and smaller mussels), crustaceans (shrimps, scuds and crayfish), and aquatic insects (mayflies, stoneflies, caddis flies etc.). They break down organic matter, regenerate nutrients and serve as food for many birds and fish (Muniz 1991). Changes in the population structures of BMIs are relatively easy to assess because they have short life cycles. Many species are acid-sensitive and they have proved useful as indicators of acidification (Økland and Økland 1986).

Studies have shown that populations of mayflies consistently decreased with increases in acidity, and when a critical level of acidity is reached they become extinct (Harriman and Morrison 1982). Raddum and Fjellheim (1984, cited in Muniz 1991) showed that in Norwegian lakes, 60% of mayfly species were lost when pH reached 5.5. However, only 30% of stonefly species were lost for the same pH decrease.

The response in amphipods to increased acidity is similar to that of the aquatic insects, mollusc and crustacean populations. However, changes in the crustacean and mollusc populations must be interpreted with care, because as they require calcium for their shells and scales a shortage in calcium may be the main determinand of mollusc and crustacean occurrence, though the calcium content and pH of surface waters are often strongly correlated (Økland 1983).

Declines in species richness and BMI diversity are closely related to decreases in pH observed in acidified lakes. The loss of sensitive species is, however, sometimes partially

offset at moderate acidity by the positive response of the more acid-tolerant species (Muniz 1991).

The acidification of surface waters can lead to distinct physiological responses in BMI populations, which affect ion and osmoregulatory processes. BMIs also demonstrate reduced growth rates when affected by acidification. This may be the result of wider food web changes leading to availability of an inferior food source: a shift in species composition due to acidification could reduce the density and diversity of prey species available (Muniz 1991). Decreased microbial decomposition of allochthonous matter should lead to increases in those species that can process coarse detrital matter (Stenson and Eriksson 1989) and Townsend *et al.* (1983) have observed shifts in shredder/scraper guilds in acidified lakes.

1.2.5 Freshwater Fish

A survey of fish stocks in the lakes of the LaCloche Mountains, Ontario, Canada confirmed the link between the presence and absence of fish species and water chemistry, in particular the acidity of the lake water (Matuszek and Beggs 1988). There was a gradual decline in the numbers of fish species with increasing acidity (Matuszek and Beggs 1988) starting with the loss of cyprinid minnows below a pH of 6.0. Similar observations have been made in populations of the Lake Trout in the Adirondacks Mountains in New York State, USA. Lake Trout is often found to be absent from lakes when the pH of the water is less than 5.4, with Brook Trout being lost when pH falls below 5.1 (Schofield and Driscoll 1987). Similar responses to acidification have been demonstrated by Almer *et al.* (1974) in Scandinavian lakes where, with increasing acidity, there is a similar sequence of disappearance; minnows first followed by the roach species and finally by the pike and eels.

Loss of fish species can be seen in decreasing fish densities and biomass, and in reduced productivity and yields of fish populations (Harriman *et al.* 1987). The rate of extinction depends upon the degree of acid stress, the life history and longevity of the different fish species, and their acid tolerance (Muniz 1991). The different life stages (egg, sac fry, emergent fry, juveniles and adults) have different sensitivities to acidity; therefore, many acid-stressed fish populations are characterised by a dominance of older fish. Younger fish are absent as a result of the partial or total failure in recruitment resulting from acid stress

during reproduction and development of the young when they are more sensitive to acid stress (Muniz 1991).

Beamish *et al.* (1975) attributed poor growth in acid-stressed fish populations to the reduction in prey, themselves affected by acidification. This is particularly apparent in piscivirous species like pike or Lake Trout. Mills *et al.* (1987) found that when minnows disappeared from Lake 223, Experimental Lakes Area, Ontario, Canada, Lake Trout became emaciated because of the reduced food supply.

Disruption of ion and/or osmoregulation in gills is the main physiological response to toxic exposure from lowered pH and high concentrations of aqueous aluminium (McWilliams and Potts 1978). This can lead to a loss of body salts and to secondary effects on the water and ion balances of the internal tissues and structures of fish. In low pH conditions, ion regulatory effects are primary, but at slightly higher pH and calcium levels the toxic effects of aluminium cause respiratory stresses and are more important (Muniz 1991). Aluminium precipitated on to the surfaces of the gills effects the diffusion of gases across the gill lamellae (Ultsch and Gros 1979), causing changes to the structure of the gills (Leino, Wilkinson, and Anderson 1987; Evans, Brown, and Hara 1988), and changes in the enzyme process of chloride cells which are involved in the uptake of salt in the gill epithelia (Muniz 1991). It appears, however, that the calcium concentration in lake water plays an important part in determining the effect of low pH conditions on the gill membranes and respiratory functions of fish (Howells 1995). In low calcium-concentration waters, the permeability of the gill membrane is increased leading to an amplified loss of plasma electrolytes. Fish have been found in highly acid waters, with little or no ill effects due to pH, where calcium or other ion concentrations have remained high (Howells 1995).

1.2.6 Birds

Evidence of direct, toxic effects on waterfowl that are the result of acidification has not been substantiated (Muniz 1991). Indirectly, however, waterfowl may be adversely affected by the loss of other organisms upon which the waterfowl prey. Many acid-sensitive BMIs are eliminated from waters that are still suitable for fish populations. This may result in predatory fish and birds competing for a declining food source; both becoming

impoverished (DesGranges and Hunter 1987; McAuley and Longcore 1988). Waterfowl may thrive, however, if predatory fish are absent or become absent via the effects of acid deposition (Hunter *et al.* 1986). Acidification leads to greater water clarity, and this can favour those bird species that use visual cues to hunt (Eriksson 1984). Reductions in the crustacean and insect larval populations, which result from the effects of acid deposition, may, however, lead to a reduction in the size and quality of surviving food sources (Blancher and McNicol 1988).

The performance of the dipper, *Cinclus cinclus*, has been related to the loss of acid-sensitive prey: mayflies and caddis fly larvae from breeding ponds (Ormerod *et al.* 1986; Ormerod *et al.* 1991). Dippers usually select sites with abundant invertebrates. Along the River Ifron, dipper populations decreased between 1960 and 1984 as acidity increased by 1.7 pH units (Ormerod *et al.* 1991).

1.2.7 Amphibians

Some widespread amphibian species are conspicuous by their absence from low pH and/or high aluminium ponds (Albers and Prouty 1987). Many species' habitats for reproduction are vulnerable to acidification (Pough 1976; Pough and Wilson 1977). This is exacerbated because the spawning period in early spring coincides with acid pulse events that can be particularly toxic to embryonic and larval stages.

In the UK, Beebee *et al.* (1990) documented the decline of the natterjack toad, *Bufo calamita*, which used to breed in ponds in Woolmer Forest that are now too acid to allow the development of embryonic and larval stages. Evidence of diatoms, macrophytes, and heavy metal and SCP deposits in the sediments of ponds in Woolmer Forest suggest that the declines in amphibian populations were associated with acidic deposition (Flower *et al.* 1988; Beebee *et al.* 1990).

1.2.8 Microbes

Bacteria, protozoa and fungi process organic compounds that regenerate nutrients used by other biota, they are also a source of food for many benthic organisms. It was originally

thought that acidification caused a shift from bacteria-dominated to fungus-dominated microbial communities (Bick and Drews 1973). This has been questioned more recently, with the evidence suggesting that bacterioplankton communities decrease in size if the acidification causes a reduction in the concentration of dissolved organic carbon in the lake water (Muniz 1991).

Microbial abundance appears to be closely linked to the processes of decomposition and nutrient cycling. Acidification can cause accumulations of leaves, twigs and other detrital allochthonous matter, and this has been used as an indicator of decreased decomposition by the microbial community (Grahn, Hultberg, and Kandner 1974). Minshall & Minshall, (1978) observed decreased litter decomposition rates in acidified streams and the littoral areas of acid lakes. Liming can lead to recovery of microbial communities as decomposition rates increased when lime was added to an acidified lake (Hultberg and Andersson 1982).

1.3 Palaeolimnology

Palaeolimnology is the study of the history of lakes. It is concerned with how lakes have changed over time and with understanding those changes, primarily through an assessment of the historical record contained in the sediments of lakes.

Over the last few decades numerous sampling and analytical techniques have been developed that enable palaeolimnologists to obtain sediment cores containing undisturbed sediment records (e.g. Glew 1991; Charles, Smol, and Engstrom 1994; Glew, Smol, and Last 2001). These can then be analysed for a variety of fossil organisms and geochemical markers using appropriate laboratory methods (Berglund 1986; Last and Smol 2001a; Smol, Birks, and Last 2001a; Last and Smol 2001b; Smol, Birks, and Last 2001b). Dating techniques using ^{210}Pb (Appleby *et al.* 1986) and other radiometric methodologies (Charles *et al.* 1994) can be used to generate chronologies for the sediment record. This allows the timing and rate of change to be determined.

Sediments found within lakes are derived from three main sources; atmospheric inputs, catchment inputs and lake inputs (Figure 1). Atmospheric sources include wet and dry deposition both from sources within the catchment and those from outside.

Inflows from streams and groundwater contributions make up the catchment component (allochthonous material), and the planktonic, benthic and littoral flora and fauna of the lake contribute to the lake input fraction (autochthonous material) (Battarbee *et al.* 1988a).

Palaeolimnological studies are often based upon studies of the fossil remains of biological indicators, the aquatic flora and fauna living in a lake. Sediment records of change are averaged over time, the species composition representative of the predominant or average hydrochemical or physical conditions of the lake. Short-term cycles, such as diurnal or diel changes, can influence the water chemistry or the biological sample taken during spot sampling measures (Anderson and Battarbee 1994). However, the historical record of lake sediments is time averaged, and, therefore, does not contain the degree of noise inherent in spot sampling records.

Organisms that are preserved in the sediment of a lake, such as diatoms or Cladocera, can be analysed for changes in community structure over time. Assessment of the modern distribution of these organisms can be related to the hydrochemistry of the water body, and optima and tolerance ranges for individual species can be calculated (Battarbee 1991). These optima and tolerance ranges can be applied to the fossil communities by means of transfer functions (Birks, Juggins, and Line 1990a; Birks *et al.* 1990b; Battarbee 1991; Charles *et al.* 1994; Dixit and Smol 1994; Bennion 1994). These enable changes in the hydrochemistry of a lake to be determined from the fossil community assemblages deposited to the sediment at that time (Anderson 1995).

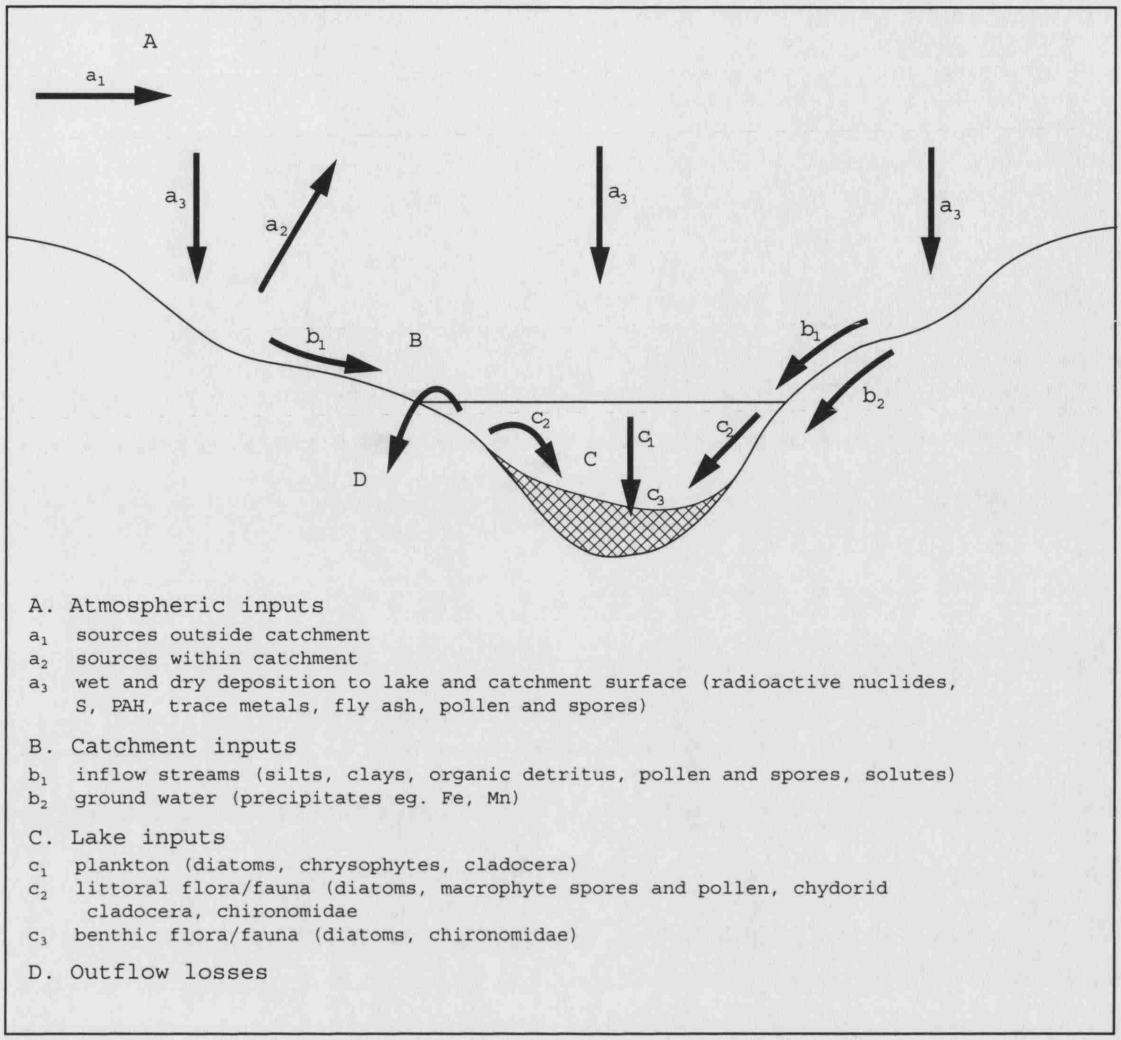


Figure 1: Schematic cross-section of a lake and its catchment showing sources and pathways of material found in lake sediments (from Battarbee et al. 1988a).

These techniques are now widely used within palaeolimnology. They have been used to determine changes in water bodies related to many hydrochemical variables, including acidity (pH) (Birks *et al.* 1990b; Dixit, Dixit, and Smol 1992a; Dixit, Dixit, and Smol 1992b; Cumming *et al.* 1994; Allott *et al.* 1995), total phosphorus (TP) (Dixit and Smol 1994; Bennion 1994), Dissolved Organic Carbon (DOC) (Kingston and Birks 1990), salinity (Fritz *et al.* 1990; Fritz 1990), aluminium (Al) (Dixit *et al.* 1992b), and more recently they have been used to infer changes in climate and temperature (Lotter *et al.* 1997).

Analyses of the sediment record for contamination by pollution indicators such as heavy metals (Pb, Zn etc.) (Davis *et al.* 1983), spheroidal carbonaceous particles (SCP) (Wik, Renberg, and Darley 1986; Rose *et al.* 1994; Rose, Golding, and Battarbee 1996) and inorganic ash spheres (IAS) (Rose 1996), have allowed the causal mechanisms for pollution

induced changes to be determined, especially in the case of the acute acidification of sensitive surface waters in Northern Europe and North America.

Palaeolimnological techniques have been used to assess changes in both the biology the chemistry of lakes. The techniques used in palaeolimnology have also been used in monitoring projects attempting to evaluate current changes in aquatic systems. Dixit and Smol (1994), present four transfer functions that can be used to reconstruct the hydrochemistry of lakes from the enumeration of siliceous microfossils from lake sediments. In this way, inferences about the current, changing, hydrochemical nature of surface waters can be made using the biological record of organisms contained within recent lake sediments.

The short instrumental time series available from contemporary studies are often unsuitable for identifying trends (Smol 1995). Natural variability is difficult or impossible to distinguish from data of insufficient quality or time span. Without long-term data, environmental managers often have difficulty in determining the trajectory or the causes of degradation, let alone the likely effects of recovery or targets for their mitagatory efforts (Smol 1995). Smol (1990) argues that the task of assessing ecosystem health is made harder without suitable and available long-term data. Palaeolimnological data provides such long-term environmental, hydrochemical, lithological and biological information, enabling the reference state (pre-impact) of surface waters to be established (see section 1.3.1 below). Comparisons between present day and the reference states of surface waters provide invaluable information on the degree of change present at a site. Where anthropogenic pollution is suspected or known to have degraded surface waters, targets for the restoration of the site can be established with regard to these reference conditions.

The wide variety of organisms preserved in the sediments and the range of chemical, lithological and biological assays that can be applied to the sediment allows for a wide range of experiments and hypotheses to be tested using palaeolimnological techniques.

1.3.1 Sub-fossil remains found in lake sediments

Palaeolimnological techniques are dependent upon the preservation of organisms in the sediments of lake systems. Not all conditions are suitable for the preservation of

organisms and the sub-fossils found in lakes can be dissolved or broken (Flower 1993; Cameron 1995). Some groups of organisms are not preserved within the sediment record. Preservation can also be dependant on the environmental characteristics of the lake. For example, ostracods are not found in surface waters that have low calcium concentrations because in these conditions their ability to grow a strong calcite shell is impaired and they are less protected against predation. However, a number of important species groups are reliably preserved in the sediments of low alkalinity lakes. It is upon these organisms that the majority of the development of palaeolimnological techniques has taken place. These groups are described below with examples of their use in palaeolimnological studies.

1.3.1.1 Cladocera

Cladocera, commonly known as 'water fleas' on account of their jerky swimming motion, are a group of microscopic crustaceans. Their true position in the phylogenetic tree is Kingdom Animalia, Phylum Arthropoda, Sub phylum Crustacea, Class Branchiopoda. Branchiopods are a group of distantly related orders of crustaceans. The orders have few characteristics in common, principally flattened thoracic legs (phyllopods) and mandibles. The phyllopods are flat, generally unbranched and non-segmented appendages that are edged with setae. This characteristic appears to be the result of convergent evolution rather than commonality of ancestry. The mandibles of branchiopods are simple rods with inner grinding surfaces. Branchiopods also have a pair of spines or claws on the last body segment though the evolutionary significance of these structures is unclear (Dodson and Frey 1991).

Previous attempts at classifying the Cladocera suggested that the group was multi-phyletic and divided them into four orders (Dodson and Frey 1991). Recent molecular evidence indicates that the Cladocera are actually a monophyletic group and have been reorganised into the single order Cladocera (Latreille, 1829), and the four old orders have been used as suborders. The four suborders of the order Cladocera are Anomopoda, Ctenopoda, Onychopoda and Haplopoda. These four suborders are divided into 11 families, approximately 80 genera and about 400 species (Korhola and Rautio 2001). Table 2 shows the orders and families of Cladocera and approximate number of genera per family.

Cladocera have been found in almost all forms of aquatic habitat from large lakes to ponds. They have been found in ditches, puddles, caves, saturated moss beds and even in moss growing on trees in rain forest several metres off the ground (Frey 1988). The Cladocera are an extremely important member of the fauna of aquatic systems performing a variety of functions by grazing on algae, detritus and heterotrophs, by recycling nutrients to lower trophic levels, by acting as a source of food for higher level predators.

Table 2: Orders, families and number of genera of the Cladocera from the class Branchiopoda. The numbers of genera are only approximations as the group is under constant taxonomic revision. (Source Dodson and Frey 1991)

Sub Order	Family	Genera	Number of Genera
Anomopoda	Daphniidae	*Ceriodaphnia, Daphnia, Dahpniopsis, Megafenestra, Simocephalus, Scapholeberis*	6
	Moinidae	*Moina, Moinodaphnia*	2
	Bosminidae	*Bosmina, Bosminopsis*	2
	Macrothricidae	*Acantholeberis, Bunops, Drepanothrix, Echinisca, Grimaldina, Guernella, Iheringhula, Ilyocryptus, Lathonura, Macrothrix, Neothrix, Ophryoxus, Parophryoxus*	17
	Chydoridae	*Acroperus, Alona, Alonella, Alonopsis, Anchistropus, Archepleuroxus, Australochydorus, Biapertura, Bryospilus, Camptocercus, Celsinotum, Chydorus, Dadaya, Disparalona, Dunhevedia, Ephemeroporus, Euryalona, Eurycercus, Graptoleberis, Kurzia, Leberis, Leydigia, Monope, Monopsilus, Notoalona, Oxyurella, Paralona, Phrixura, Planicirclus, Pleuroxus, Plurispina, Pseudochydorus, Rak, Rhynchotalona, Saycia, Spinalona, Tretacephala*	32
Ctenopoda	Sididae	*Diaphanosoma, Latona, Latonopsis, Penilia, Pseudosida, Sarsilatona, Sida*	8
	Holopediidae	*Holopedium*	1
Onychopoda	Polyphemidae	*Polyphemus*	1
	Cercopagidae	*Bythotrephes, Cercopagis*	2
	Podonidae	*Caspievadne, Cornigerius, Evadne, Pleopsis, Podon, Podonevadne*	1
Haplopoda	Leptodoridae	*Leptodora*	1

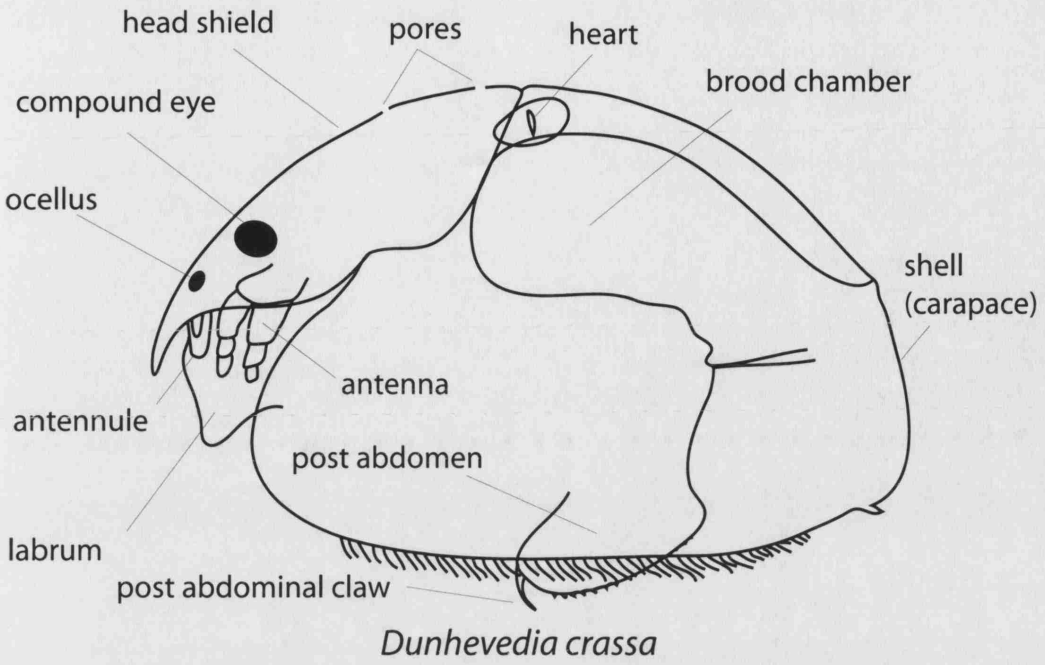


Figure 2: Diagram of Dunhevedia crassa KING, 1853 (Cladocera, Chydoridae) *showing the main features of the Cladocera. Redrawn from course notes by C Duigan.*

Cladocera have a single compound eye and a carapace that is used as a brood chamber. The carapace generally envelops the body of the cladoceran, though in the suborders Onychopoda and Haplopoda this is not the case as the carapace only encloses the brood chamber. The head of the cladoceran smoothly joins the body, which hangs within the carapace, with the carapace and the head being connected at the back of the neck. The carapace is often ornamented with striae, geometric patterns, spines, teeth, depressions or a hexagonal mesh.

The swimming limbs, the second antennae, are attached to the body either side of the head, and partly, or completely, protrude from the carapace. The second antennae are large branched appendages and are used to propel the cladoceran about the water column. The thorax of the Cladocera bears four to six pairs of legs that beat rhythmically. The legs have a flattened leaf-like appearance with the coxopodite bearing a flattened exopodite of gill. The paired limbs perform the function of respiration, but they are also used for filter feeding. Food collected on the legs is passed towards the head of the cladoceran by the rhythmic beating of the limbs. The mouth is located near to the margin that separates the head from the carapace. A large mandible is found on either side of the mouth, which consists of a single rod with the inner chewing surfaces found immediately anterior to the mouth.

The first antennae (antennules) are located near the front of the head. These are paired appendages, generally short in length, that are thought to be sensory organs. In the genus *Moina* and in many of the members of the Macrothricidae, however, these appendages can be longer than the head. In *Bosmina*, the antennules have developed to form a long tusk-like structure.

The abdomen of the Cladocera bears a pair of claws at its tip. These claws groom the thoracic legs when the abdomen is swung forward. Some smaller members of the Cladocera, especially the Chydoridae, also use the foot in a kicking action to move over the various surfaces that they inhabit, accommodating their limited ability for swimming (Fryer 1968). The tip of the abdomen, where the claws are found, is known as the postabdomen as it is located ventral to the anus.

In most females, the space between the body and the carapace is used as the brood chamber. In the members of the Onychopoda and Haplopoda the body is external to the carapace and the carapace is used entirely as a brood chamber in these cases, located on the back of the animal.

Throughout the growing season the majority, if not all, the Cladocera present in a lake will be females. Reproduction is by unfertilised female-bearing eggs. These eggs are nurtured in the brood chamber. When conditions deteriorate, either climatically or the result of overcrowding or changes in the availability of food supplies, males are produced parthenogenetically. Resting eggs are generally fertilised by the males, and the embryo undergoes several cycles of cell division before it enters diapause. The resting eggs are protected by a modification of the carapace, the ephippium, and are resistant to freezing and drying. The ephippium is thicker than the normal carapace and is very darkly pigmented. The ephippium is shed when the female molts and the edges of the ephippium close around the resting egg. In general, the rest of the carapace separates from the ephippium, though in *Daphnia* the tail spine, or mucro, is also incorporated in the ephippium. The resting eggs require conditions to become more favourable before they recommence development. These eggs generally produce females who then quickly reproduce asexually to develop a viable population (Dodson and Frey 1991).

Cladocera are represented in lake sediments by a variety of body parts. The remains commonly found are of the headshield, carapace, postabdomen, postabdominal claws, and ephippia. These remains differ in their morphology between species and based on these differences it is possible to differentiate between species. Cladoceran remains recovered from lake sediments are primarily found to be derived from the littoral-benthic dwelling family Chydoridae. Remains of planktonic Cladocera are generally restricted to the genus *Bosmina*, although remains of *Daphnia* (especially the ephippia and postabdominal claws) are routinely recovered (Frey 1960a; Frey 1962a).

Sedimentary records of Cladocera have previously been used to show the development from oligotrophy to eutrophy in lakes (Nilssen and Sandøy 1990). Clear changes in the species composition of the cladoceran assemblages, where there is distinct replacement of species within the Chydoridae and *Bosmina*, can be related to changes in fish predation pressure and lake trophy.

The direct response of the Cladocera to acidification is less clear. Interpreting changes in species assemblages and abundances over small changes in pH is complex, and cladoceran autecology is poorly understood below pH 5.5. Changes in species composition often related to acidification could reflect indirect effects, resulting from changes in predation pressure and vegetation type and cover that are seen in aquatic systems undergoing acidification.

A number of palaeolimnological studies however, have demonstrated changes in cladoceran assemblages related to acidification (Nilssen and Sandøy 1990; Paterson 1994; Stemberger and Lazorchek 1994; Post, Frost, and Kitchell 1995; Hann and Turner 2000; Tremel *et al.* 2000; Walseng and Karlsen 2001). Increasing acidity has been demonstrated to effect changes in community interactions, the loss of acid-sensitive species and of species richness as a whole, and changes in total individual numbers of Cladocera.

1.3.1.2 Diatoms (Bacillariophyceae)

Diatoms are single celled, golden-brown algae characterised by an external siliceous cell wall, or frustule, in which silicic acid has been dehydrated and polymerised to form silica particles (Wetzel 1983). The frustule consists of two valves connected by bands of silica

known as girdle bands. The frustule is often highly ornamented with various species exhibiting different patterning. It is this ornamentation on the valve and its general shape and nature that form the basis of taxonomic separation of diatom species in many floras (Barber and Haworth 1981). Cell wall morphology varies between genera and species, and even allows varieties and forms of species to be distinguished.

Diatoms are found almost everywhere that light and moisture occur, including virtually all marine, brackish and freshwater environments, as well as soils, ice, and attached to rocks and other substrates within spray and splash zones near water. Diatoms live singly or form colonies, usually secreting a mucilaginous material that covers the frustule and allows the diatom cells to attach to one another or to the substrate (Wetzel 1983). Diatoms live on a variety of benthic substrates, as well as in the plankton of lakes.

The diatom plankton can be divided into the euplankton, the meroplankton and the tychoplankton. Euplanktonic diatoms spend their entire life suspended in the water column and whilst common in the marine environment, many planktonic diatoms in lakes have adapted to employ a resting stage that allows them to weather unfavourable conditions. In these cases they are considered meroplanktonic. Tychoplanktonic taxa are those that are commonly found in the benthos but which have become resuspended into the water column (Battarbee *et al.* 2001).

Benthic diatoms are associated with habitats located around the margins of lakes; on rocks and stones (the epilithon), on aquatic macrophytes (epiphyton) on mosses (epibryon), on sand grains (epipsammon), and sediments (epipelon). Their extension into deeper habitats is dependent upon light attenuation and suitability of habitats. The epilithon and epiphyton communities have many species in common, but the longevity of the habitat itself often limits of long-term community development, e.g. die back of aquatic macrophytes during winter (Battarbee *et al.* 2001). The episammic community is very distinct, with taxa needing to be capable of surviving periods of exposure to dark or anoxic conditions. Epipelic taxa are adapted to low light conditions and to avoid burial in sediments, being motile and therefore able to move through interstitial waters of the sediment-water interface.

Due to the siliceous nature of the frustule, it is resistant to a certain degree of chemical attack and is usually well preserved within lake sediments. The relative ease with which diatom samples can be taken and analysed, and the variety and beauty of the various forms of frustule, have resulted in diatoms being a long-studied member of the limnic flora (Battarbee and Charles 1986). Diatom autecology has largely been described in the literature and observations regarding tolerances and optimal abundances for a variety of hydrochemical variables have been made (See section 1.3.2).

The key controls on diatom community composition in lakes are mainly physical and chemical in nature. The availability of light, temperature and turbulence are important physical factors, whilst pH, nutrients and salinity are the key chemical factors. pH is perhaps the most important factor controlling diatom species composition but there is little detailed ecophysiological understanding of how pH leads to physiological stress and ultimately to determine diatom species composition (Battarbee *et al.* 2001).

Sections 1.3.2 and 1.3.3 discuss the role of diatoms in palaeoecological studies of lake water pH and acidification.

1.3.2 Hydrochemical Reconstruction using Palaeolimnology

It has long been recognised that by studying the autecology of biological indicators for various hydrochemical variables it is possible to classify specific groups of organisms that are found together under similar hydrochemical conditions. Conversely, the hydrochemistry of surface waters can be inferred through an analysis of the organisms living in them. In recent years sophisticated statistical techniques have been used to provide more accurate, robust hydrochemical reconstructions from fossil records that have a firm ecological basis.

Early approaches to pH reconstruction have been made qualitatively using Hustedt's pH classification system using diatoms. Nygaard (1956, cited in Battarbee and Charles 1986) further enhanced this approach by developing a number of indices. These indices were based on ratios of the percentages of diatom valves in Hustedt's pH categories. Meriläinen (1967) further developed quantitative approaches for reconstructing the acidity of surface

waters using the relationship between the $\log_{10}$ of index values (e.g. Index α) and the measurements of lake water pH using regression analysis. The slope and intercept of the regression equation were then used to predict lake pH. Renberg and Hellberg (1982) derived Index B, again based on pH categories, which was also used to predict pH. Index B uses more information than Index α, and is less reliant on alkaline taxa that are rare or absent in acid lakes. Multiple linear regression of the optima and tolerance ranges of biological indicators for pH has also been used successfully to reconstruct the historical record of pH change from sedimentary records (Flower 1986).

The Surface Waters Acidification Project (SWAP) (Battarbee and Renberg 1990) and the Palaeoecological Investigation of Recent Lake Acidification (PIRLA) (Charles and Whitehead 1986) projects developed new methods of reconstructing hydrochemical variables from the diatom species assemblages found within the sediment record. The technique developed for SWAP and PIRLA (Birks *et al.* 1990a; e.g. Birks *et al.* 1990b) was based around weighted averaging (WA) regression and calibration (ter Braak 1987). Weighted averaging regression and calibration overcome many of the problems of other calibration methods (Korsman and Birks 1996). WA assumes a unimodal relationship for the response of species to explanatory variables. This is considered to be a sound ecological assumption (ter Braak and Prentice 1988). Other regression and calibration methods assume linear responses to explanatory variables and, therefore, do not represent ecological responses of species to environmental gradients as well as WA. WA also maximises the covariance between the species data and the measured environmental variables (Korsman and Birks 1996). This is the same approach used in direct gradient analyses such as canonical correspondence analysis (CCA). Indirect gradient methods attempt to maximise the variance only within the species data and, consequently, some information may be lost when only the first few components are used for regression. If more components are used, multi-colinearity problems are introduced to the analysis (Korsman and Birks 1996).

There are some weaknesses in WA, but these have recently been addressed with the development of a new variation of WA; Weighted Averaging Partial Least Squares (PLS) regression (ter Braak *et al.* 1993; ter Braak and Juggins 1993). WA-PLS uses the residual structure of the species data within the regression and calibration procedure to improve the predictions made using the technique (ter Braak and Juggins 1993). Simple WA fails to

accommodate this extra information and the predictive power of the WA calibration models produced using this method may not be as accurate as those developed using WA-PLS.

The calibration approach involves a two-step process. Firstly the relationship between the species data and the measured environmental variables (the predictor variables) is established using WA or WA-PLS regression. The relationships derived from the regression procedure are then regressed or calibrated against the fossil data to predict pH from the fossil assemblages using inverse WA or WA-PLS regression.

To test the transfer function model, the predicted results are compared to a set of observed data. Strictly, the model should be tested against independent data, not against the data from which the model was derived. However, Birks *et al.* (1990b) demonstrated the techniques of 'Bootstrapping' or 'leave-one-out jack-knifing' that estimate the true error of the model. These methods achieve this by taking a sub-sample of the training set to compare with the observed data, thus forming an independent test of a model's predictive power. These techniques are computer intensive, with *ca.* 1000 calibrations being run to test the error of prediction in a transfer function. Bootstrapping is especially suitable as the whole dataset is used to test the model.

1.3.3 Palaeolimnology and lake acidification

Palaeolimnology played an essential role in the study of lake acidification. During the early 1980s, there was considerable debate surrounding the cause of fish stock declines in upland lakes. Many scientists, especially those from Scandinavia and Canada where the effects of acid deposition were first described, believed that acid emissions were to blame for the recent acidification of surface waters in Europe and North America (e.g. Odén 1968). A number of alternative hypotheses were formulated, however, in what was a highly charged scientific and political debate.

One such claim was that acid lakes were the result of long-term, natural acidification processes (Pennington 1984). It was claimed, that after lakes were formed when the glaciers retreated and climate warmed at the start of the Holocene, lakes acidified as weathering of soils led to the progressive leaching of base cations. Soils would gradually

acidify and the subsequent runoff would slowly lower the pH of lakes. The basis behind this hypothesis was that there are many lakes that today are presently acid but that have no historical record of fish stocks, evidence for long-term acidification. Battarbee (1984) reviewed the evidence for long-term acidification as the cause of acute surface water acidification in the present day. The decline in pH identified in the sediment record of certain lakes is very slow (often less than 0.1 pH unit per 1000 years, much slower than the rate of acidification recorded in many presently acidified lakes (Battarbee 1990).

Other claims surrounded the large-scale changes in land-use within upland areas of Norway (Rosenqvist 1978). Decreased grazing in lake catchments may have led to an increase in heathland vegetation for example and with it an increase in acid soils. Evidence from Sweden, for example, (Renberg, Korsman, and Anderson 1993a; Renberg, Korsman, and Birks 1993b) has shown considerable influence on lake hydrochemistry by changes in land-use. Many lakes in southern Sweden were shown, by hydrochemical reconstruction using a diatom-pH transfer function, to have an alkaline phase prior to the onset of acidification in the 19th Century. The alkaline phase was the result of land-use practices that no longer exist in modern day Sweden. The extent of the recent acidification of many such Swedish lakes is much greater than that which would be expected if land-use changes alone were the cause. In the last few decades many of these lakes have become much more acidic than at any point throughout their history. (Renberg *et al.* 1993a; Renberg *et al.* 1993b)

More credibility was placed upon claims that recent acidification of surface waters was the result of recent afforestation. Streams draining from afforested catchments have been shown to be more acidic than those draining non-afforested areas (e.g. Harriman and Morrison 1982). Many of the afforested regions of the UK are located in areas of high sensitivity to acid deposition with base-poor bedrock, slow weathering rates and high rainfall (Kinniburgh and Edmunds 1986). Consequently, many areas of afforested upland Britain have low acid buffering capacities. Tree growth processes (Nilsson, Miller, and Miller 1982; Nilsson 1993), enhanced scavenging and foliar uptake of sulphur dioxide (Lindberg and Garten 1988), as well as land improvement measures prior to tree planting (Hornung and Newson 1986), were all proposed as mechanisms which promoted acidification.

Palaeolimnological studies have demonstrated that, whilst afforestation can result in enhanced acidification, it cannot account for the widespread, rapid acidification of European surface waters. Kreiser *et al.* (1990) studied four Scottish lochs, two with afforested catchments (Loch Chon and Loch Doilet) and two with moorland catchments (Loch Tinker and Lochan Dubh). Loch Tinker and Loch Chon are located in the Trossachs region of Scotland, an area receiving high levels of acid deposition, whereas Loch Doilet and Lochan Dubh are located in an area that receives low levels of acid deposition. Kreiser *et al.* (1990) demonstrated that acidification occurred gradually from *ca.* 1850 at Loch Chon. When the site was afforested in the 1950s, there was an increase in the rate of acidification. The most rapid period of change occurs in the 1960s following canopy closure when pH fell from 5.8 pH units to 5.2. Loch Tinker shows some signs of an early decline in the planktonic *Cyclotella* flora of the lake, a biological indicator of the onset of acidification. There was further acidification up to the 1930s, but the acute, recent (post 1960) acidification and associated diatom changes shown to occur in Loch Chon did not occur in Loch Tinker. There was no overall change in inferred pH or diatom composition above a depth of 8cm (1930) in the sediment core. pH reconstruction at Loch Tinker showed that pH fell from 6.6 to 5.7 by 1930, but then fluctuated around 5.6-5.7 until the present day. Carbonaceous particle records from the two lochs indicated that both were receiving high levels of atmospheric contamination by the 1940s, which is prior to the afforestation of Loch Tinker. In contrast the diatom sub-fossil flora of Loch Doilet, the other Loch in the study with an afforested catchment, did not indicate any further acidification following afforestation. Loch Doilet receives lower levels of deposition than Loch Chon and the forest has had little or no effect of the chemistry of Loch Doilet. The two Lochs with moorland catchments both acidified after *ca.* 1850, precluding afforestation as the primary cause of surface water acidification.

Palaeolimnological techniques were also used to evaluate the claims that acidification of surface waters was the result of deposition of fossil fuel derived acids to lakes and the catchments draining into them (Battarbee *et al.* 1985; Battarbee *et al.* 1988a). Recent lake acidification is strongly correlated with evidence of atmospheric contamination both from trace metals and SCPs from lake sediments (Battarbee 1990). Records of carbonaceous particles document the impact of atmospheric contamination to surface waters (Wik *et al.* 1986; Rose 1995; Rose 1996). Created during high-temperature burning of fossil fuels in power stations, carbonaceous particles first enter the sediment record during the nineteenth

century as the industrial revolution took hold in the UK. Carbonaceous particle concentrations drastically increase in the post-war period (1950s), reaching a peak around 1970. Concentrations decline after the 1970s because of international efforts to curb emissions in the late 1970s and early 1980s and the introduction of electrostatic precipitators on industrial chimney stacks.

Battarbee (1990) demonstrated that the trend in surface water acidification seen in UK lakes paralleled the trends in carbonaceous particle concentration identified from the sediment record. The first indications of acidification occurred around *ca.*1850 and rarely before 1800 in UK lakes. This is also the period where a response of the diatom flora to the contamination is first observed (e.g. Battarbee *et al.* 1988b). Contamination levels increase rapidly after *ca.* 1940, and the diatom flora shows a rapid change to more acid tolerant species after this time. Lakes that have little evidence of acidification in the sediment record, such as Loch Corrie nan Arr (Battarbee 1990), are also those that have not acidified despite their sensitivity. On the other hand, in those sensitive areas that receive high loadings of atmospheric deposition acute acidification of the surface waters has been identified from the sediment record (Battarbee 1990).

Palaeolimnology has played a crucial role in determining the causes of lake acidification. Without the high resolution historical record contained in the sediment of lake basins and interpreted by palaeolimnologists it would have been particularly difficult or impossible to prove the role emissions from power stations played in acidifying surface waters.

Since SWAP and PIRLA, the emphasis has shifted towards identifying chemical and biological recovery from acidification. Dixit *et al.* (1989) showed clear recovery in Swan Lake using chrysophytes. Swan Lake, situated near to the Sudbury smelters in Ontario, had acidified because of the large amount of SO_2 and toxic metal pollution emitted from the smelters. Dixit *et al.* (1992b) demonstrated similar recovery in three other Sudbury lakes, this time using both diatom and chrysophyte remains. Recovery was attributed to the reduction in emissions from the Sudbury smelters where emissions of SO_2 had fallen by 50%.

Evidence of recovery in sites remote from point source pollution was described at the Round Loch of Glenhead (Battarbee *et al.* 1988b; Allott, Harriman, and Battarbee 1992).

In cores with high sediment accumulation rates recent changes in the diatom flora of the loch were identified, including an increase in the abundances of several acid sensitive taxa. These studies have indicated that recovery from acidification should happen in lake systems following emissions reductions, though detecting this may be more difficult in sites remote from point-source emissions.

1.4 Restoration of acidified lakes

1.4.1 Concepts of ecological restoration

Ecosystems can be related to each other in terms of their structure and their function. The number of species and the organisational complexity of the ecosystem define ecosystem structure. Ecosystem function is a combination of the biomass and nutrient content of that ecosystem (Bradshaw 1984). Degradation of an ecosystem will result in the reduction of ecosystem function or structure or both. Bradshaw (1984) represented this in the form of a diagram (Figure 3). Natural ecosystem processes will move a degraded ecosystem along a theoretical pathway back to its original state. This process will take a long time to complete and can only take place if the forcing that led to the degradation has been removed. If this has not been removed, further degradation of the ecosystem may take place. Moving the ecosystem along the pathway to the original state artificially is the act of restoration.

Bradshaw (1984; 1996) proposed a number of options for use in situations where restoration may not be possible; the original state of the ecosystem may be the result of human land use practices that are now outdated and no longer practised, for example. The options then are rehabilitation and replacement. Rehabilitation is progress made towards the original state that is not complete. Ecosystem function and structure have been improved, but the pre-disturbance conditions have not yet been achieved. Alternatively, another ecosystem can be substituted for the degraded one. The substitute ecosystem is generally less complex than the original state. This process is known as replacement. True restoration may be unrealistic in many situations. It may be prohibitively expensive or inappropriate. In these circumstances, rehabilitation or replacement may be better suited to providing an ecosystem more valuable in function and complexity than the degraded ecosystem.

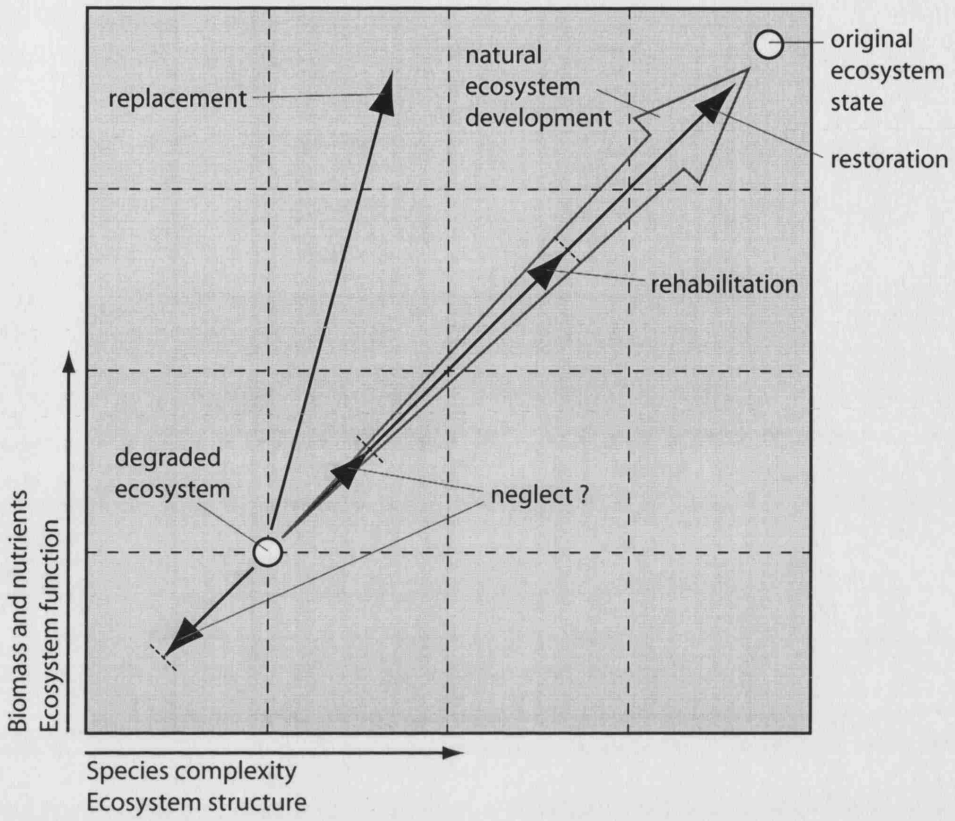


Figure 3: Changes in ecosystem structure and function for a hypothetical degraded ecosystem. Note the various forms of 'restoration' and the changes associated with true restoration (redrawn from Bradshaw, 1984).

Bradshaw (1996) suggested that by aiming at ecosystem restoration we are setting too high a target for ourselves. The term 'ecosystem' describes both the biological and the non-biological elements that occur together in a given area. Restoration of an 'ecosystem' then should revolve around restoring the function, structure, and the interaction of the whole system. Consequently, the current emphasis in restoration is on 'habitat' restoration where the importance is placed upon restoring the 'place' where organisms live (the habitat) rather than the processes of a degraded ecosystem.

Good ecological restoration entails negotiating the best possible outcome for a specific site based on ecological knowledge and the diverse perspectives of interested stakeholders (Higgs 1997). It should be noted then, that simply restoring a disturbed ecosystem to its former state does not always form the most appropriate use of the available funds or the most appropriate restored community.

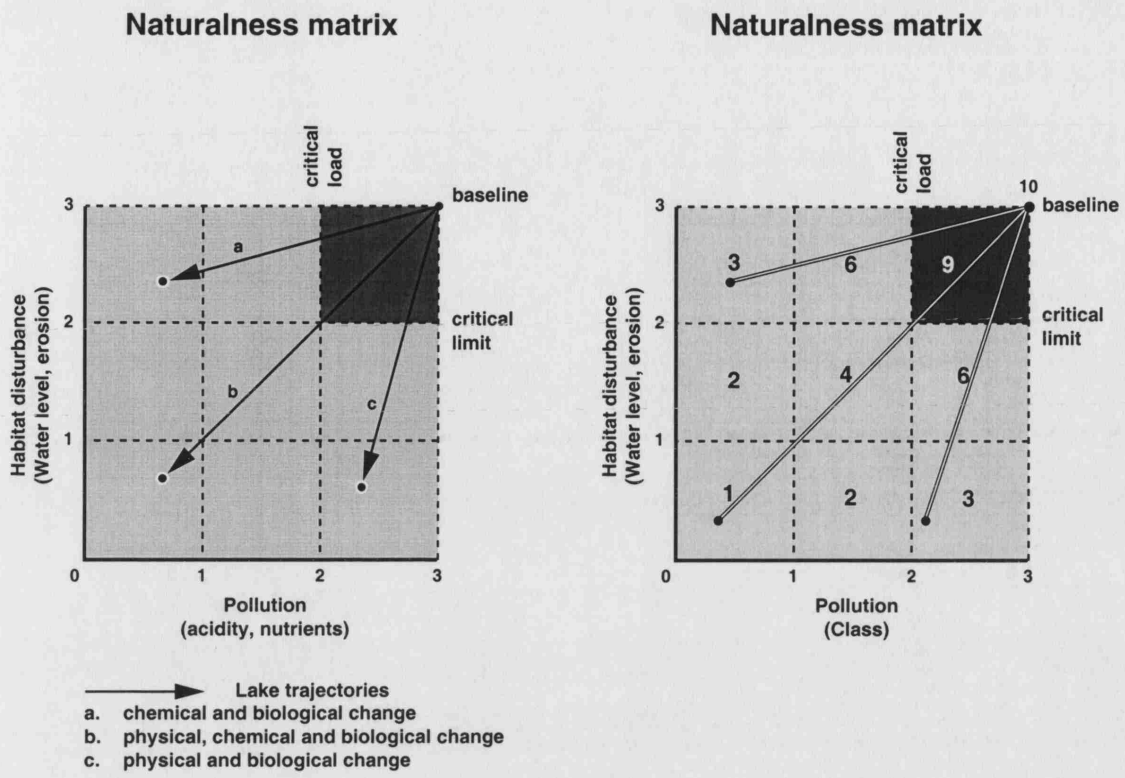


Figure 4: Naturalness matrix showing: (1) hypothetical change over baseline for three lakes (a, b, & c): (2) a 'state-changed' classification system related to change over baseline, with 1=most changed, 9=least changed and 10=pristine (from Battarbee 1997).

1.4.2 Lake restoration

The challenge of restoring lakes has been a driving force behind limnological and palaeolimnological research for the last twenty years or so. The combination of limnoecological work and palaeolimnological studies of pre-disturbance conditions in degraded lake systems is a potentially powerful approach to tackling the issues surrounding the restoration of lake ecosystems.

From an ecological standpoint restoration of aquatic ecosystems should represent the restoration of biological activity: achieving working ecosystems in which macrophytes, zooplankton, plankton and other aquatic fauna are functioning within their normal range of activity.

Freshwaters are perturbed by two different sets of impacts (Battarbee 1997); contamination of the surface waters from a diverse range of pollutants and habitat disturbances such as physical alterations to the shoreline or catchment (e.g. hydroelectric dam construction). How pollution and habitat disturbance interact to determine the nature of environmental change in freshwaters needs to be considered if the restoration of freshwaters is to be achieved.

Battarbee (1997) illustrated the relationship between pollution effects and habitat disturbance in the form of a naturalness matrix (Figure 4). The pollution axis refers to chemical change that may cause a biological change; acidification and eutrophication being the two factors of primary importance in lake systems (Battarbee 1997; Battarbee 1999). Other pollutants include contamination by toxic metals persistent organic pollutants (POPs) and radionuclides. Biological change is seen when the pollution load exceeds the *critical load* for that pollutant. The habitat axis is defined as ecologically important disturbances such as shoreline alteration, water-level regulation, inwash of catchment soils and direct biological disturbances, such as over fishing or the introduction of new species. Again a threshold can be defined, after which, significant biological change will take place, and in the case of habitat disturbance, this threshold is termed the *critical limit* (Battarbee 1997).

Numbers can be applied to the squares of the natural matrix that reflect the degree of disturbance in terms of habitat and pollution. A score of 10 would indicate pristine conditions with now change over baseline conditions being observed. A score of 9 indicates sites that are minimally impacted, where some physical or chemical change has taken place, but those changes have not exceeded the critical load or limit of a lake and therefore have led to little identifiable biological change (Battarbee 1997). Battarbee (1997) has argued that this is perhaps the highest attainable target for lakes in the UK and that this should be the main target of water management. The other scores reflect the degree of biological change, with scores being lowest where both extensive physical and chemical changes have taken place. Battarbee provides examples of lakes where disturbances fall into particular squares on the matrix (Battarbee 1997; Battarbee 1999).

1.4.2.1 The problem of eutrophication

Restoration practices in lakes have mostly centred on the problem of cultural eutrophication and the reversal of the environmental damage caused by the enrichment of surface waters. In eutrophicated surface waters, the disturbance is mainly one of an addition of nutrients (nitrogen [N] and phosphorous [P]) either directly to the lake from the catchment or indirectly via rivers and streams that flow into the lake. The pollution may be either from diffuse sources (e.g. leachate and runoff from agricultural land) or point sources (e.g. sewage effluent discharges).

The addition of nutrients to a lake leads to an increase in lake productivity. This increase in productivity brings with it a whole host of changes to both the water column and the biota supported by the lake.

Bottom-up approaches to mitigating the impacts of eutrophication have involved reducing inputs to lakes to reduce the pollutant load and manipulation of the hypolimnion through aeration, whilst top-down approaches, such as biomanipulation of fish stocks and the implantation of artificial refugia have also been attempted (Moss *et al.* 1996; Lindenschmidt and Hamblin 1997; Bootsma, Barendregt, and van Alphen 1999; Little *et al.* 2000; Søndergaard *et al.* 2000; Brouwer, Bobbink, and Roelofs 2002; Søndergaard *et al.* 2002).

These techniques have been applied successfully in many lakes in Europe and beyond. They are however, of limited use in acidified lakes. The counterpart in acidified systems to the diversion of the source of nutrients is the reduction of the emissions of sulphur and nitrogen. This is process is being taken forward, from the point of view of the UK, by discussions within the European Union (see section 1.1.1). The main alternative interventionist approach that can be applied to acidified lakes is the process of liming lakes and their catchments to increase the acid-buffering capacity of the water and soils.

1.4.2.2 Liming: an example of lake acidification restoration

One technique employed by those attempting to restore acidified surface waters has been to lime either the water directly or the catchment that drains into the acidified lake (Renberg and Hultberg 1992). Although the impetus behind such restoration methods has been a biological one it is still essentially a hydrochemical approach in nature. The aim of

liming is to elevate lake-water pH (and increase acid neutralising capacity (ANC) and decrease Al toxicity as well) so that it will allow fish populations to be re-introduced or to re-colonise.

Each year approximately 300 000 tons of powdered limestone is used to treat lakes, rivers and wetlands in order to increase the pH of surface waters across the globe (Henrikson, Hindar, and Thornelof 1995). This costs between 40-50 million US$ per year (Henrikson *et al.* 1995). There are over 11 000 lakes that are routinely limed in Sweden and Norway alone.

Originally, the main aim of liming was to restore fish stocks to impacted lakes for economic and recreational fishing (Henrikson *et al.* 1995). These aims have now been broadened to encompass the preservation and recovery of biodiversity and human health. The official aims of the liming projects in Sweden (Svenson *et al.* 1995) and Norway (Romundstad and Sandøy 1995) are to allow for the preservation and recolonisation of flora and fauna, to raise the pH of surface waters and to improve the conditions for recreational fishing.

In the UK there has been far less use of liming to mitigate the impacts of acid deposition (c.f. Howells 1995). At Loch Fleet, Southwest Scotland, an extensive study of liming applications and the biological and chemical responses to the treatment began in 1986. Around 450 tonnes of $CaCO_3$ was applied to various parts of the catchment, and the water quality of the loch quickly improved following the treatment where the pH rose to 6-7, calcium to 4 mg l^{-1} and labile Al was greatly reduced (Howells 1995).

The water quality proved suitable for the return of trout through restocking, and 18 months after the initial treatment the restocked fish had shown good growth and fecundity (op cit.). There were also changes in the invertebrate fauna of Loch Fleet, a decline in predatory beetles (thought to be due to fish predation) and the return of some acid-sensitive species. Palaeolimnological studies of Loch Fleet were conducted (Flower *et al.* 1990; Cameron 1995) using diatoms to determine the biological response of this important group to the lime treatment. They concluded that the pH of the lake had risen to a level that had never been reached throughout the history of the lake (Anderson *et al.* 1986; Flower *et al.* 1990). The diatom community present after liming contained species that

were previously rare or absent from the diatom sedimentary record (e.g. *Synedra arcus*). Other examples from the literature include (Renberg and Hultberg 1992; Howells, Dalziel, and Turnpenny 1992; Renberg *et al.* 1993a; Henrikson *et al.* 1995; Nilssen and Wærvågen 2002)

This demonstrates the crux of the problem with liming as a restoration tool. Far from restoring the pre-acidification flora and fauna of acidified lakes the results show that conditions created in lakes after liming are often different to any condition previously seen in the lakes throughout the palaeolimnological record. This cannot be considered restoration in the sense of Bradshaw rather it is replacement. In essence, a degraded ecosystem is being replaced by one that restores a target organism to the treated lake but that does not reflect the original state of the lake in terms of structure or function. This may be an appropriate target for which to aim if the biological target is the restoration of fish stocks to impacted lakes. However, it should not be taken to represent the original state of the degraded ecosystem. For true restoration to be achieved the biology of the original state of the degraded ecosystem must be restored. Identifying such targets or reference states, therefore, should involve biological as well as hydrochemical variables and should be related to conditions that are similar to those previously exhibited during the history of any given lake. Palaeolimnology can define such targets or reference states in terms of the hydrochemistry and selected biological indicators.

1.4.3 The use of baseline conditions as restoration targets or reference states

Most surface waters in Scotland, Wales and Northern England were formed as glaciers retreated when the climate began to warm at the end of the last glacial. During this period natural climatological, geological, geomorphological and biological change has taken place as well as progressive anthropogenic impacts since the Neolithic period. More recently, increasing anthropogenic pressures have begun to affect UK surface waters, most notably the eutrophication and acidification of lakes and rivers. Catchment processes or the hydrochemistry of certain lakes may have buffered some or all of the effects of change. Other, more sensitive, surface waters may have been altered by the natural and anthropogenic changes to which they have been exposed.

The potential influence of natural factors has clear implications when we consider defining the baseline state of degraded (acidified) surface waters. The reference state of the surface water could be defined as either the pristine state for a surface water or the state prior to the onset of specific perturbation. The potential effects of climate change on lake water acidity is discussed in more detail below (see section 1.6.4). Such natural responses complicate the process of defining baseline state for acidified lakes, let alone describing the hydrochemical and biological status of such a state.

Bradshaw (1984) used the term '*original state*' to define the pre-impact state of an ecosystem. In palaeolimnology, this state can be described as the reference state of surface waters. The reference state of surface waters is that which combines an absence of habitat disturbance and an absence of pollution. It can be argued that this state represents the highest quality of surface water that is attainable in the UK, and that these conditions should be the main goal of any restoration project.

The reference can also be thought of as the state prior to a specific impact. In the UK, evidence of acidification is rarely found before the 1850s. This period can be defined as the reference state for acidification but does not necessarily reflect the state where there has been no anthropogenic influence on the lake. In this way reference states can be used to define future hydrochemical and biological reference conditions, or if unobtainable, due to irreversible change, alternative targets that approximate natural conditions can be defined (Battarbee 1997). In such cases, some of the most significant baseline biological characteristics can be defined from the sediment record contained within the lake.

1.4.4 Hydrochemical and Biological Targets for Restoration

Studies of disturbance in lakes often concern the change in concentrations of a certain chemical variable or suite of associated variables in the lake water, for example, total phosphorus (TP) (e.g. Bennion 1994), pH (e.g. Flower and Battarbee 1983) or Al and DOC (e.g. Kingston and Birks 1990), or a biological change such as the loss of fish populations or the formation of algal blooms. Consequently, targets could be set with reference to either or both hydrochemical and biological criteria. Reconstructing the hydrochemical characteristics of surface waters has often been achieved through the use of indicator species (Birks *et al.* 1990a; Birks *et al.* 1990b); the hydrochemical change being defined by

the response of a particular group of organisms to that change. Consequently, because of this need to quantify biological change in response to environmental change for hydrochemical reconstruction, the biological response to acidification is fairly well understood. The effects of lake acidification have long been identified, with the major loss of fish populations and aquatic vegetation within acidified systems being the most immediately identifiable of responses of the system to decreasing pH (Uutala 1990; Kingston *et al.* 1992; Havas and Rosseland 1995). Other studies have looked at the effects on algal populations, such as diatoms or chrysophytes (Battarbee 1984), whilst other authors have described the change in insect (Brodin and Gransberg 1993) or zooplankton (Havas and Rosseland 1995) populations associated with acidification.

The restoration of fish stocks to lakes is an easily recognisable, ecological (biological) target for scientists, politicians and the environmentally aware public to focus upon, as well as being an issue of great economic importance. The reliance upon this sort of target, and the use of solely hydrochemical parameters to define it, has slowed progress towards adopting relevant biological reference states that address the whole lake system rather than a small part of a lake's ecosystem. Henriksen *et al.* (1990) sampled many lakes across Norway for the 1986 1000-lakes survey. The project was designed to determine the surface water chemistry and the status of fish populations of lakes from acid sensitive areas of Norway. The data were used to identify changes in water chemistry and fish status since 1974 and 1975 (c.f. Henriksen *et al.* 1990) and to act as baseline characteristics against which the effects of emission reductions (e.g. LRTAP protocols) could be judged. The only biological aspect incorporated in the 1000-lakes survey was the fish status of the surveyed lakes. It was only later, through further study of the other biological groups found in aquatic systems, that the degree to which other organisms were affected by acidification was identified (Havas & Rosseland, 1995).

1.4.5 Defining baseline states and restoration targets

To define a baseline or reference state for a surface water, hydrochemical reconstructions have traditionally been used to provide information about the reference state hydrochemistry of surface waters. There are two potential ways that hydrochemical reconstructions can be made; dynamic modelling and transfer functions. Dynamic modelling uses information about the hydrological and chemical processes in a catchment

including variables such as weathering rates, soil CO_2 partial pressures, and cation exchange capacity, to model hydrochemical responses to given deposition patterns. One such dynamic model is MAGIC (Model of Acidification of Groundwaters in Catchments) (Cosby *et al.* 1985a; Cosby *et al.* 1985b; Jenkins *et al.* 1990; Jenkins, Ferrier, and Cosby 1997; Cosby *et al.* 2001). Using known estimated deposition levels the models have be used to hindcast the historical hydrochemistry of surface waters. However, most often the models are used to predict future hydrochemical targets for given levels of emission abatement such as the LRTAP conventions discussed above (see section 1.1.1). The other main way in which hydrochemical reconstructions are made is using palaeolimnological techniques and hydrochemical transfer function (see Section 1.3.2).

1.5 Biological Restoration Targets and the Modern Analogue Approach

Reference states can be defined chemically, biologically, or both. The chemical approach uses transfer functions or hydrochemical models. However, a biological approach to defining biological targets or reference states using palaeolimnological data is the modern analogue approach. The modern analogue approach uses robust statistical methods to compare fossil assemblages to modern assemblages. This approach was developed by Flower *et al.* (1997).

1.5.1 Modern Analogues

"Modern Analogues" are lakes that have modern chemistries and biological communities that match the past conditions of a specific lake. Defining analogue sites enables the biological characteristics of the analogue lakes to be used to define biological targets or reference states for recovery from acidification. Space-for-time substitution is the basis of selecting sites as modern analogues. This assumes that the aquatic communities currently existing in analogue sites closely resemble those that existed in similar but now acidified habitats (Flower, Juggins, and Battarbee 1997). However, the sites and the biological communities they contain can vary greatly through subtle differences in geology, climate or other biogeographic and chemical variables.

The methodology behind analogue matching in palaeoecology has its foundations in the field of palynology (Overpeck, Webb, and Prentice 1985). Fossil pollen spectra derived from palaeoecological studies were compared to the pollen spectra retrieved from pollen traps and taken from modern sites, allowing modern forest ecosystems to be matched with those that existed in the past. To generate a fuller understanding of the past inferences could be made about ecosystem structure and function from the modern analogues. Developments in computing an numerical techniques have led to the increasing use of statistical methods to quantify the degree to which modern sites match the fossil flora (c.f. Overpeck *et al.* 1985). These statistical methods are known as dissimilarity coefficients.

Dissimilarity coefficients measure the difference between multivariate (e.g. species abundances) samples such as those produced in palaeolimnological studies. Dissimilarity coefficients have a number of advantages: they are a precise method of comparison between samples and the process may be automated or computerised. Dissimilarity coefficients allow the opportunity to calibrate the dissimilarity scale in terms of the underlying biological or environmental differences. Other multivariate approaches also exist (e.g. using principal components analysis or canonical variates analysis) but these do not directly compare the degree of analogy that dissimilarity coefficients measure.

Prentice (1980) identified three types of dissimilarity coefficient, the simple or unweighted, the equal weight, and the signal-to-noise measures. Simple methods tend to be heavily influenced by taxa that have wide ranges, and the rare types have little influence upon the results. Equal weight measures up-weight the rare taxa and down-weight the common species. A potential problem is that these tests may give extra weight to the insignificant, yet potentially noisy, taxa. The method of up-weighting is dependent upon the statistical test being used. Signal-to-noise measures avoid this by weighting the data so that the signal component of the difference between fossil and modern samples is emphasised at the expense of the noise component. The problem with signal-to-noise measures is that they are scale dependent (Webb, Laseski, and Bernabo 1978; Delcourt, Delcourt, and Webb 1983). Equal weight measures may prove more appropriate if certain minor taxa are judged important to the analysis. Overpeck *et al.* (1985) tested various dissimilarity coefficients for application in analogue matching tests. They found that the signal-to-noise methods provided the more robust analysis because rare taxa are down-weighted and the

floristic signal drawn out. Their results also showed that the three signal-to-noise measures they tested all provided similar results.

There has been very limited use of modern analogues, especially in any quantitative way, within the field of palaeolimnology to tackle the issues of recovery in acidified lake systems. An exception is the work of Flower *et al.* (1997), who identified modern analogues for two Scottish lochs from a modern dataset of lakes from northern Europe. Flower *et al.* (1997) used the squared chi-square distance measure in their study of diatom-based modern analogues for the pre-acidification status of recently acidified lakes in northern Europe. They compared the fossil diatom assemblages in two acidified lakes from the Galloway region of Scotland (the Round Loch of Glenhead and Loch Dee) with modern surface sediment samples from over 200 lakes in Britain, Ireland, Sweden and Norway. The squared chi-square analysis identified several modern analogue sites within the dataset. Loch Teanga in the Hebrides and Lough Claggan, Ireland had the most similar diatom floras to the pre-acidification floras of the Round Loch of Glenhead and Loch Dee respectively. These sites, however, can not be considered as being pristine sites (i.e. sites that have not been impacted by atmospheric deposition). Loch Teanga and Lough Claggan are located in areas of low to moderate acid deposition (0.4-0.8 g S m^{-2} yr^{-1}) (Battarbee *et al.* 1988a; Flower *et al.* 1994; Flower *et al.* 1997). Although the diatom communities of these lakes do not show any response to this level of deposition other biological communities may have been affected by deposition or trace metal contamination. Flower *et al.* (1997) also looked at the other hydrochemical determinands of the lakes selected as analogue matches for the Round Loch of Glenhead and Loch Dee. The Round Loch of Glenhead had no modern analogues that had a similar lake-water calcium concentration. Loch Dee, however, had some analogues with similar calcium concentrations (e.g. Loch Howie, Scotland, and Lough Brockagh, Donegal) and these sites may represent better analogue sites than Lough Claggan.

1.6 Key Assumptions and Problems of the Modern Analogue Approach

The modern analogue approach makes a number of assumptions about the data used and the theoretical concepts of analogue matching (c.f. Flower *et al.* 1997). These assumptions need to be tested before analogue matching techniques are applied to the study of recovery,

as they are inherent in understanding the validity of substituting space for time in biological studies.

The key assumptions of the modern analogue approach are: (1) the fossil group (or groups) used for analogue matching is representative of the wider range of biological groups found in fresh waters, (2) the modern surface sediment dataset contains suitable analogue lakes for the acidified lakes that will be matched with them (e.g. that there are enough pristine or minimally impacted, sites to provide an adequate range of matches) and (3) the hydrochemistry and biological communities of analogue sites accurately represent the pre-acidification conditions of acidified lakes.

1.6.1 Representation of the wider biological communities in fresh waters by diatoms.

It has to be assumed that the fossil group used in analogue matching (e.g. diatoms in the Flower *et al.*, (1997) approach) is representative of the whole trophic structure and variation in other biological groups found in aquatic systems (e.g. other algal groups, aquatic macrophytes, other invertebrate groups, or fish). Furthermore, the variation in the fossil record is assumed to reflect accurately changes occurring to the whole biology of a given lake. This has to be assumed because not all groups leave remains that are preserved in the sediment record of lakes (see section 1.3). Therefore, analysing change in other biotic groups is impossible without these inferences unless historical records are available.

It is well known that diatoms respond quickly and with definite pattern to changing lake-water acidity. Given this information, it is likely that a matching process based solely on sub-fossil diatom data is selecting analogues with similar pH characteristics to those inferred from the surface sediments of presently acidified lakes rather than on the whole hydrochemical signature. It is likely, therefore, to expect that the results do not necessarily reflect whole ecosystem characteristics where the concentrations important determinands, such as labile aluminium, calcium or DOC, can vary considerably at a given pH.

Section 1.2 above and the references therein have shown that the biological response to acidification is not simply a response to changing pH and H^+ ion concentrations for many species groups. In many cases, increasing concentrations of aluminium, which becomes

more biologically available as lake water acidity increases, lead to direct toxic effects on many species groups, notably fish and zooplankton. Calcium concentrations also play a role in determining the biological response to acidification. Calcium plays a role in mitigating the effects of acidity and Al toxicity by regulating the permeability of the gill membrane in fish. High calcium concentrations can lead to a reduction in the loss of plasma electrolytes across the gill membrane. Whilst diatoms are good indicators of Al and DOC concentrations in lake water (Kingston and Birks 1990; Birks *et al.* 1990a; Birks *et al.* 1990b) they are poor indicators of calcium concentrations (c.f. Flower *et al.* 1997).

Due to differences in the degree of preservation in lake sediments amongst the biological groups found in fresh waters the choice of proxies for use in analogue matching is quite restricted. Practical considerations must also be taken into account when choosing a proxy, such at the amount of sediment required to screen for a specific proxy. The macrofossil remains of aquatic macrophytes (for example, seeds and fruits) can provide a great deal of information on changes in aquatic macrophyte communities through time (Smol *et al.* 2001a), yet require large amounts of sediment to achieve a representative sample and the best coring location is not in the deepest point of a lake basin, unlike the majority of other proxies used in palaeolimnological investigations.

Cladocera are a suitable group worthy of further investigation for use in an analogue matching procedure because they are a ubiquitous group, well represented in the sediments of oligotrophic lakes. The remains of Cladocera found in sediments contain representatives of the planktonic, the littoral benthic and the epiphytic cladoceran communities, and community composition is determined by, and related to, a wide range of hydrochemical, physical and biological factors (e.g. Quade 1969; Fryer 1980; Krause-Dellin and Steinberg 1984; Fryer 1985; Duigan 1992; Korhola 1999).

The Cladocera are complementary to the diatoms because of the positions the two groups occupy in the trophic structure of lake ecosystems. Diatoms are amongst the many primary producers in lakes and occupy a wide range of habitat types. Cladocera are generally herbivorous grazers feeding on benthic, epiphytic and planktonic phytoplankton. Cladocera themselves are prey for predatory zooplankters, invertebrates and fish. As such Cladocera play an important role in the transfer of energy from the primary producers up to higher levels in the trophic structure of a lake.

1.6.2 Coverage of sites with varying hydrochemical characteristics in the modern training set.

In the Flower *et al.* (1997) application, the lakes selected as the closest analogues to the Round Loch of Glenhead and Loch Dee were lakes that had been somewhat impacted by acid deposition. They were from areas receiving low to moderate acid deposition. Flower *et al.* (1997) suggested that this was due to there being insufficient pristine sites in their dataset. Another problem they encountered was that the two selected analogue lakes had very different hydrochemistries apart from their pH levels (as inferred from the diatom data). Consequently, it is important to assume that the dataset being used for analogue matching has sufficient geographical range and hydrochemical scope for the matching process to be worthwhile.

Flower *et al.* (1997) recognised the need to incorporate surface sediment data from more pristine sites. In their dataset, there were very few pristine sites, the majority being found in northern Norway. Including more pristine sites could improve the chances of pristine lakes being selected as modern analogues. This is problematic because there are very few areas of Northern Europe that can be considered pristine as there are large areas of northern Europe that have been impacted by acid deposition.

1.6.3 Representing the hydrochemistry and biological communities of pre-acidification conditions of acidified lakes from those of modern analogue sites.

For the process of space-for-time substitution to be accepted as an adequate model, it must be assumed that the hydrochemistry and biology of analogue lakes are a true reflection of pre-acidification conditions observed in acid lakes. This assumes that anthropogenic acidification is the only or the major forcing mechanism affecting environmental change upon acidified lakes in the UK over the Holocene period. This also assumes that climate change or other anthropogenic influence (e.g. habitat disturbance) since pre-acidification times (*ca.* AD 1850) has had negligible effects upon the overall hydrochemical and the biological functioning of the lake systems.

There is some degree of evidence that climate changes and other anthropogenic influences (e.g. catchment disturbance or land-use change) have had profound impacts on the

hydrochemistry and biology of some studied lakes across northern Europe (e.g. Renberg *et al.* 1993a; Renberg *et al.* 1993b).

1.6.4 Variability in surface water systems

Nature is inherently dynamic: a constantly changing or fluctuating climate has been a major driving force in determining the present day distribution of ecosystems and the physical appearance of the landscape that they occupy. Among this background of climate change and natural variability anthropogenic activities have influenced and changed ecosystems, predominately since the industrial revolution, but also during the forest clearances since the Neolithic period (Roberts 1989).

An important consideration in the ecological restoration of acidified sites is the degree to which the pre-acidification hydrochemical, and ecological and underlying habitat conditions may have changed. There are three main themes that need addressing when considering the fluctuating hydrochemical and ecological conditions of surface waters, the climate effect, catchment disturbance and land-use change effects, and the apparent stability of UK lakes prior to anthropogenic acidification.

1.6.4.1 Climate

For mountain lakes, Skjelkvåle & Wright (1998) have argued that palaeolimnological analogues may be of little use because future climate changes are likely to cause 'conditions never previously experienced on earth, such as high atmospheric CO_2 levels and high UV-B radiation.' (Skjelkvale and Wright 1998 pp. 285). A number of other studies have also described the likely effects of a changing climate for acid lakes (e.g. Psenner and Schmidt 1992; Schindler *et al.* 1996; Leavitt *et al.* 1997). Such changes would compromise the applicability of the modern analogue approach and question whether targets for recovery from anthropogenic acidification are realistic goals.

To assess the impact or influence that fluctuations in climate have had upon lakes, long core stratigraphies that contain past records of climate fluctuation are required. Analysis of the cores for changes in biological assemblages over time and dating of these changes

can answer questions surrounding whether upland acid waters were chemically and ecologically stable prior to anthropogenic acidification or are continually changing, systems.

Psenner & Schmidt (1992) have demonstrated a relationship between colder air temperatures and lower pH of surface waters in two soft water, high-altitude lakes in the central Alps. The inferred pH data from diatom analysis and correlation with the temperature record was also supported by the results of loss-on-ignition (LOI) and Fe/Mn ratio analysis of the lake sediment. LOI and Fe/Mn are surrogates for the organic content of the sediment.

Psenner & Schmidt (1992) found three distinct peaks in the LOI and Fe/Mn data that corresponded to the temperature peaks of 1820, 1860, and 1900 in Austrian and Swiss instrumental records. Prior to the onset of anthropogenic acidification in the late 19th and early 20th centuries there was a distinct coupling between climate and biogeochemical processes in the two alpine lakes. This coupling has progressively broken down over the last 100-150 years as the deposition of strong acids to surface waters has had greater influence on the biogeochemical processes of the lakes than the fluctuating climate. This has clear implications for assessment of recovery in systems that have climate/biogeochemical coupling. They suggest that recovery observed in some systems, by a levelling off of the inferred pH decline, may be due to rising temperatures and global warming as much as that due to decreasing acid deposition (Psenner and Schmidt 1992).

Monteith and Evans (2000) have recently identified decadal fluctuations in the hydrochemistry of a number of upland lakes on the west of Britain, which appear to be linked to climate cycles and specifically to the North Atlantic Oscillation (NAO). These trends are currently being investigated further using high resolution palaeolimnological techniques at a lake in northwest Scotland (Monteith and Flower, *pers. comm*).

This section has described a number of studies demonstrating the effects of climate change on aquatic ecosystems. As discussed by Skjelkvåle & Wright (Skjelkvale and Wright 1998), if climatic variability and change both in the past and into the future is significant this poses considerable challenges for the use of modern analogues in restoration-type work. This manifests itself via uncertainty in the suitability of a single sediment sample from the pre-disturbance phase to represent natural background conditions, and also via

uncertainty in the suitability of using a modern analogue as a future recovery target in light of predicted climate change; your target may change due to one climatic driver, and this driver may not be the main driver forcing change in the acidified system for example. In effect, the end result is likely to be a shifting of the goal-posts one is aiming for, and exactly how one modifies the target to take account of this is uncertain. This, the second area requires further work and is to be tackled in an EU Framework 6 research programme, EURO-LIMPACS (Battarbee, pers. commun.).

The first is more easily dealt with. Section 1.6.4.3 below, variability of pre-disturbance phase sediment samples is discussed, and whilst there is little evidence of variability in pre-disturbance sediment samples, these have been sampled at only a course resolution. It should not be forgotten though, that the time averaging effect of sedimentation processes leads to a single pre-disturbance sample representing 5 or 10 years of sediment accumulation per cm slice of sediment. As such, these samples are an average of the conditions over that period and short term fluctuations will be smoothed. Work is ongoing to assess the variability of conditions seen in pre-disturbance sediment samples for the UK AWMN lakes (Curtis, pers. Commun.).

1.6.4.2 Catchment disturbance and land-use change

For Swedish lakes, Renberg (1990) and co-workers (Renberg *et al.* 1990; Renberg *et al.* 1993a; Renberg *et al.* 1993b) have produced long-core stratigraphies and have analysed the diatom profiles of a number of lakes. The results show features characteristic of diatom-inferred pH profiles from some of the acidified lakes in Northern and Southern Sweden. Swedish lakes are generally base rich, well buffered and mesotrophic immediately following the end of glaciation. Following this, a period of slow natural acidification takes place where progressive leaching and loss of base cations from the soils leads to soil acidification and dilution of the Acid Neutralising Capacity (ANC) of run-off entering the lakes (Renberg *et al.* 1993a; Renberg *et al.* 1993b). The pH of the Swedish lakes immediately following retreat of the ice sheets has been shown to be around 7.0 and the slow process of natural acidification increased acidity to *ca.* pH 5.5.

Acidification continued up until *ca.* 2300 BP when suddenly, and across a geographic range, the pH of lakes rose quickly to *ca.* 6.5. This has been attributed to a shift in land-use in the

catchments of many Swedish lakes known to have taken place around this period of history because of an expansion in the agrarian culture of the Iron Age. Increases in pH have been correlated with the recession of natural forest, expansion of shrub vegetation, increased frequency of cereals and weeds, increased concentrations of charcoal and LOI values, indicating increased soil erosion (Renberg *et al.* 1993a; Renberg *et al.* 1993b). Cultural alkalisation continued until the industrialisation of the 1900s. Many lakes in Sweden have acidified considerably since the 1950s; many lakes becoming severely acidified with permanently reduced pH values of between 4.5 and 5.0.

Renberg *et al.* (1993b) also identified a recent liming period in many of the acidified southern Swedish lakes. Liming has been adopted as a widespread restorative measure in over 6000 lakes (Renberg *et al.* 1993b). Liming characteristically increases pH to levels above those of the immediate post-glacial, and the resultant diatom flora is quite unlike anything that has been found in the post-glacial history of the Swedish lakes.

From this type of study, it is clear that considerable anthropogenic influences on lakes have occurred in southern Sweden since around 2300 BP. This has implications for any attempt to restore these acidified lakes to a former state. The problem now facing environmental managers is that if they are to restore lakes to pre-acidification states then they will be attempting to return lakes to conditions that were inherently dependant upon the cultural changes in land-use associated with the expansion of agriculture in the Iron Age. This type of land-use no longer exists in Sweden so there is little hope of restoring lakes to such a status. Renberg *et al.* (1993b) extrapolated the predicted pH response to continued natural acidification given that none of the changes in land-use or deposition chemistry of the last 2300 years had taken place. Arguably, this provides a suitable target that could be attainable given pollution abatement strategies already in place or being negotiated. This is a theoretical attempt to express the expected trajectory following the overriding trend in natural acidification over the previous 10000 years following deglaciation. However, this provides no more representative a target than the pre-acidification state of surface waters. Furthermore, there is no guarantee that this is a more attainable target given possible changes in the hydrochemical, physical, biological and climatological variables since 2300 BP.

1.6.4.3 Pre-acidification chemical and biological variation in UK lakes

Atkinson and Haworth (1990) undertook a similar study this time of two sites in the UK, Devoke Water in the Lake District, Cumbria, and Loch Sionascaig, Northwest Scotland. Similar trends of slow natural acidification to those found in Swedish lakes were found following de-glaciation. However, both lakes are not as sensitive to acid deposition as other upland sites in the UK and the Swedish lakes, and have suffered much less disturbance.

Devoke Water has recently (post 1850) begun to acidify to a point where there are considerable biological changes within the water body. Loch Sionascaig, on the other hand, is a well-buffered site and has a stable inferred pH profile for the past few thousand years. In neither of these two sites was a cultural alkalisation period identified.

Jones *et al.* (1986) presented a diatom-based reconstruction of pH for a Holocene sediment core from the Round Loch of Glenhead, Galloway, Scotland. The diatom stratigraphy of that core showed little change other than that associated with the acidification of the lake shortly after the end of the last glacial, and the recent, anthropogenic acidification. Birks (1996) also demonstrated the stability in the Round Loch of Glenhead using rate-of-change analysis to assess the degree of variability in diatom species throughout a Holocene sequence. The only significant period of change was found to be in the last 200 years, associated with the anthropogenic acidification of the loch (Allott *et al.* 1992), and the immediate post-glacial where base cation rich soils began to leach, thus influencing the ionic composition of run-off.

Given the relative stability of UK upland acidified sites it should be reasonable to assume that the pre-acidification status of a given lake is an attainable target for restoration measures to attempt to recreate. However, the apparent stability of UK upland surface waters may be misleading. Only a few sites have been studied in the UK and the resolution of the studies has been biased towards detecting acidification in the upper layers of the sediment record. There has been little high-resolution analysis of pre-1800AD sequences (c.f. Flower and Battarbee 1983; Battarbee *et al.* 1985; Flower *et al.* 1987; Battarbee *et al.* 1988b).

1.7 Aims

The principal aim of this thesis is to develop the analogue matching approach of Flower *et al.* (1997) to address some of the methodological issues of the technique raised through their work. In the preceding discussion (section 1.6 above) the case has been made to extend the existing diatom-based approach through the addition of a second proxy, the Cladocera. Additionally, recent sampling of a range of acid-sensitive fresh waters in the UK has increased the number of potential analogue sites that could be included in new approach. The principal aim of this thesis is, therefore, to develop a combined diatom and Cladocera training set for analogue matching and to apply the new approach to pre-acidification sub-fossil samples from the lakes in the UKAWMN. Additionally an evaluation of the analogue matching approach will be made, the suitability of various dissimilarity coefficients for use in analogue matching of acidified lakes will be assessed and the procedure for defining *close* modern analogues will be investigated.

This thesis has a number of secondary objectives, which stem from the work required to fulfil the principal aim. Firstly, the development of a diatom and Cladocera training set for use in a new analogue matching technique allows the opportunity to study the species environment relationships of the Cladocera; a group of fresh water zooplankton under-utilised in the field of palaeolimnology, especially in the UK. Part of this under-utilisation is due to the complex nature of the way in which the environmental factors known to influence cladoceran distributions are interwoven in fresh water systems, such as the interactions between hydrochemical changes associated with acidification and the effects of habitat and predator-prey relationship shifts that occur concomitantly with this hydrochemical change and how these are recorded in Cladocera species composition. Uncertainties in these areas limit the degree of confidence with which temporal shifts in cladoceran species composition in lake sediment cores can be attributed to environmental forcing factors. A further aim of this thesis is, therefore, to undertake an analysis of the Cladocera-physico-chemical relationships in a training set of acid, upland lakes in the UK.

Additional diatom data now available for an increased range of acid sensitive, upland lakes in the UK allows a re-evaluation of the environmental factors involved in controlling the distributions of diatom taxa in acid lakes. The SWAP calibration set published in 1990 (Birks *et al.* 1990a; Birks *et al.* 1990b; Stevenson *et al.* 1991), which has been used ever since

to reconstruct changes in pH in an ever increasing number of lakes, is derived from data from a number of north European countries (England, Scotland, Wales, Sweden and Norway). The range of data currently available for similar lakes in the UK is now sufficient for a comparison of the two training sets to be made. Therefore, a secondary objective of the thesis is to explore diatom-environment relationships in the UK training set and to compare this new training set with the existing SWAP calibration set. This aim is considered in Chapter 3: "Diatom distributions in acid-sensitive, upland lakes in the UK".

1.8 Structure of the thesis

The preceding sections of Chapter 1 introduce the problem of fresh water acidification, with particular emphasis made to recovery in acidified surface waters. The biological effects of acidification are presented and the role of palaeolimnology in the field of surface water acidification is introduced. Restoration of freshwater ecosystems is then discussed before methods for defining restoration targets are outlined. The modern analogue approach is then introduced and the assumptions of the approach discussed. Finally the aims of the thesis are stated.

Chapter 2 of the thesis describes the methods used throughout the remaining sections of the thesis. Justification for the use of certain methodologies over competing techniques is made where appropriate.

The diatom training sets are examined using a range of univariate and multivariate numerical techniques in Chapter 3 to elucidate the environmental factors that determine the distributions of diatom taxa in acid, upland fresh waters in the UK. The UKAWDDS training set is also compared to the SWAP calibration set and the differences between the two discussed.

Chapter 4 presents the results of the analysis of the cladoceran training set. Multivariate numerical techniques are used to assess cladoceran compositional responses to a range of underlying environmental factors. Species specific responses to a number of physico-chemical variables are modelled using generalised linear modelling techniques, whilst cluster analysis routines and linear discriminant analysis are used to investigate community level

response to the environmental data. The distinction between the factors involved at the three different levels is made.

Chapter 5 deals with the development and assessment of the analogue matching approach using diatom and Cladocera in the matching process. A discussion of the various dissimilarity coefficients is included and the suitability of a selection of coefficients examined. Methods to determine the level of similarity required for a *close* modern analogue are discussed and an alternative procedure outlined. The new combined diatom and cladoceran analogue matching approach is applied to the lakes of the UKAWMN and an evaluation of the ability of the approach to identify suitable analogues is made using secondary data sources. The main findings of the chapter are discussed.

Chapter 6 summarises the main findings of the thesis and discusses how the new approach could be used in future work. Methodological issues are discussed and a broader evaluation of the technique is made. The use of the approach in the future and what remaining work needs to be in evaluating and using the approach are discussed. Potential improvements to the approach are suggested.

Chapter 2: Site selection and methodology

2.1 Site Selection

Two datasets of lakes were generated under the work for this thesis, the 163-sample UK Acid Waters diatom dataset, and the 83-lake diatom and Cladocera analogue matching dataset. The UK Acid Waters Diatom Data Set (UKAWDDS) is a collection of lakes that have been sampled over the last 20 years as part of ongoing research in acid sensitive upland areas of the UK by the Environmental Change Research Centre. The basis of the training set are the UK sites included in the SWAP calibration set (Stevenson *et al.* 1991). Additional sites were then added to this data set from an assessment of acidification in lochs from northwest Scotland (Allott *et al.* 1995), a study on the effects of recent water quality changes on populations of the black throated diver (Allott and Rose 1994) and a lake classification project in Wales (Allott and Monteith 1999). A few remaining sites were included from unpublished work (ECRC, unpublished data). The UKAWDDS represents the extent of the palaeolimnological data available from acid sensitive lakes in the UK undertaken by the ECRC up to the year 2000.

The 83-lake combined diatom and Cladocera training set is formed from a subset of the samples in the UKAWDDS. Of the 163 samples included in the UKAWDDS only 83 of the surface sediment samples from which the diatom samples of the UKAWDDS were prepared had sufficient sediment remaining for a cladoceran analysis preparation (see section 2.2.2 below). This was of particular importance with some of the older sediment samples from the SWAP calibration set, which had been used for a wider range of analyses than many of the newer samples. This is shown in the geographical spread of the lakes in each of the two data sets Figure 5. Areas of the UK such as the Galloway region of southern Scotland, and the Snowdonia region of north Wales are under represented in the 83-lake training set when compared to the number of sites from these regions included in the UKAWDDS.

Figure 5: Maps showing the locations of the 163 sites in the UKAWDDS (a) and the 83-lake combined diatom and Cladocera training set (b).

Conversely, the spread of surface waters from north, northwest and northeast Scotland and mid and south Wales are equally well represented in the combined training set and the larger UKAWDDS.

Figure 5 shows maps of the locations of the sites in both the UKAWDDS and the 83-lake combined diatom and Cladocera training set. A list of sites, names, grid references for the two training sets is given in the appendices ate the end of the thesis.

2.2 Laboratory Methods

2.2.1 Diatoms

The diatom data used in the current analyses were not prepared by the author. However, standard techniques for the quantitative preparations of diatom sample were used following the methodologies of SWAP (Stevenson *et al.* 1987; Stevenson *et al.* 1991). Sediment samples were collected, usually from the deepest point in each lake, using modified Kajak, Hongve, or mini-Mackereth corers (Stevenson *et al.* 1991) or a modified Glew corere (Allott *et al.* 1995). Further details of the sampling are given in the cited references.

A number of studies in the literature attest to the close representation of lake diatom communities in sediment cores taken from an appropriate coring location (usually the deepest point of a lake) (Anderson 1986; Anderson 1989; Anderson 1990a; Anderson 1990b; Anderson 1990c; Charles *et al.* 1991; Cameron 1995). For example Cameron (Cameron 1995) compared contemporary diatom samples from Loch Fleet taken over a 20 month period during which the loch was limed. He then compared the diatom samples with sediment trap and sediment core samples and showed that the sediment core samples accurately reflected the changes seen in the contemporary diatom community resulting from the liming experiment, despite the fact that the magnitude of the change suggested by the sediment core samples was slightly muted compared to that seen in the contemporary sampling. These studies demonstrate that the use of a single sediment sample for restoration-type work is likely to be appropriate for restoration-type work.

For the Cladocera, the case is somewhat clearer, as a number of studies have shown that for the *Bosmina* and Chydoridae, the use of sediment samples taken from a suitable coring location provide a better description of the species present in a lake than repeated contemporary sampling throughout the year (Frey 1960b; Frey 1988; Korhola and Rautio 2001). This does not extend to the remaining groups of Cladocera, such as *Daphnia*, whose remains are not well preserved in lake sediments. It is appropriate therefore to draw conclusions about the *Bosmina* and chydorid communities from sediment samples, but care must be taken when interpreting the planktonic cladoceran record.

Slides of the resulting preparation were made and counts of the remains of diatoms were made by various members of the Environmental Change Research Centre (UCL) until at least 500 remains were identified.

2.2.2 Cladocera

The Cladocera samples were prepared using the standard deflocculation of 1 cubic centimetre of sediment in 100 ml hot potassium hydroxide (KOH) with gentle manual agitation for approximately 30 minutes. The method used is similar that that advocated by Frey (1986), however, many of the samples were high in organic material tightly bound to the cladoceran remains. Longer deflocculation periods were required to separate the cladoceran remains from the sediment matrix in these cases. After deflocculation, the solution containing the cladoceran remains was passed over a 33 μm sieve. The residue on the sieve was wash under a gentle flow of water to remove as much of the fin inorganic matter as possible.

The residue collected on the sieve was transferred to a 10 ml container. This was allowed to settle out overnight and the supernatant was pipetted off to remove as much water as possible. The samples were then stored in methanol with a few drops of safranin stain. Known amounts of a sample were plated onto six cover slips and mounted on slides using a small amount of glycerol jelly.

All cladoceran remains on a slide were enumerated. Slides were counted until at least 200 remains had been found. In a few cases the samples were sufficiently rich in remains to necessitate the counting of only two cover slips to encapsulate the diversity and range of Cladocera in the sample. Most samples were less rich in cladoceran exuviae and six cover slips were counted before 200 exuviae were encountered. A few samples were particularly poor in cladoceran remains and very few exuviae were counted. In these cases less than 200 remains were counted.

In these samples, the remains that were counted were spread evenly across the six cover slips. In addition, the total number of species found in these samples was reached after counting a few exuviae. It would be safe to conclude that the lakes from which these

samples were taken are species poor, supporting limited cladoceran populations and not an artefact of the sample preparation procedure.

This impression is further enhanced by the occurrence of non-separated exuviae on the cover slips. Generally, the exuviae of Cladocera encountered in sediments are individual remains of the headshield, shell, postabdomen, postabdominal claws, ephippia, anntenule segments or other remains. Sometimes where the conditions of preservation are suitable the intact remains of Cladocera can be found. If the preparatory technique was destructive, leading to the dissolution of remains, then it is unlikely that intact cladoceran remains would be encountered. Intact Cladocera were found in a number of samples including those exhibiting low concentrations of exuviae and numbers of species.

Exuviae were identified using the standard texts, including Frey (1959; 1960b; 1962a; 1962b; 1965), Goulden and Frey (Goulden and Frey 1963) and Alonso (1996). Taxonomy follows that of Flößner (1972). Taxonomic descriptions of the main taxa found in the sediment samples of the 83-lake combined diatom and cladoceran training set are given in Section 2.3 below.

Following the method of Frey (1986) all the remains attributable to one species were enumerated, and the type of exuviae that corresponded to the largest number of individuals was taken as the number of individuals for that species in the sample. The count data for the Cladocera samples were converted in numbers per cubic centimetre of sediment and percentage abundance data. These have been used for the analyses described below.

2.3 Taxonomic descriptions of selected species of Cladocera in the training set

The following section provides descriptions of the remains of the main species as identified from the sediment samples in the modern cladoceran training set. Reference is made to descriptions in the scientific literature to provide justification of the identification of certain taxa on the basis of morphological variation or to provide examples that contribute to the discussion contained below.

2.3.1 The *Bosminidae*

The taxonomy of this genus is perhaps one of the most difficult aspects of the use of cladoceran microfossils for palaeolimnological studies. A significant degree of morphological plasticity is exhibited by the species of this genus. Three species are readily recorded in the sediments of lakes, and are usually attributable to *Bosmina coregoni*, *Bosmina longispina* and *Bosmina longirostris*. Remains from these taxa are predominately the shell and the headshield, though postabdomens and claws were found in some of the sediments studied.

The headshields of the various species are generally identified from the positions and morphology of the headpores. In the *Bosminidae* the head pores are not located centrally on the headshield, but are present laterally. Goulden and Frey (1963) distinguish between *B. coregoni* and *B. longirostris* on the basis of the arrangement of these lateral headpores.

2.3.1.1 Bosmina coregoni *(BAIRD, 1857)*

The headshields of this species are distinguished by the characteristic circular pore surrounded by a pear-shaped chitinous thickening. Goulden and Frey (1963) note that the exact position of the lateral headpores varies slightly amongst specimens from different geographical areas and in the different morphological varieties of the species. In a later study Korinek (1971) remarks that in *B. coregoni* the lateral headpores generally exhibited the pear-shaped thickening though that it was not equally pronounced in all populations studied. The lateral pores are generally surrounded by characteristic reticulation. The shells of *B. coregoni* are generally large with a sizeable, often curved, mucro. The shells are also reticulated. *B. coregoni* was found in nearly all of the samples studied, often contributing up to 50% of the sample total.

2.3.1.2 Bosmina longispina *(LEYDIG, 1860)*

Headshields generally attributed to this species are remarkably similar to *B. coregoni*. Korinek (1971) indicates that headpores are situated as in *B. coregoni*, though the distinction between the two species appears to be the patterning of the reticulae around the headpores. The shells of this species tend to be small and have an enlarged, straight, pointy mucro, often up to 50-100% of the size of the shell itself.

2.3.1.3 Bosmina longirostris *(O. F. MÜLLER, 1785)*

The headshields of B. longirostris are similar in many respects to those of *B. coregoni.* The lateral headpores of this species however, are located along the margin of the antennal articulation, not located some distance anterior to the mandibular articulation as they are in *B coregoni.* The pore is oblong or rectangular and is not surrounded by a thickening of the chitin. Shells of *B. longirostris* have a very small mucro and are generally smaller than those of *B. coregoni. B. longirostris* was found only rarely in the sediments studied.

It is easy to distinguish between headshields of *B. longirostris* and those of c.f. *B. coregoni* based on these lateral headpores. Distinguishing between *B. coregoni* and *B. longispina,* however, is more difficult. The remains of *B. longispina* tend to be smaller than those of *B. coregoni,* though this is insufficient to split these species because it is unknown whether the remains are from lower instars, and as a result, a wide degree of variation in the size of the remains is exhibited.

Distinguishing between these species based on their shells is extremely difficult at present. The species *B. coregoni* exhibits a wide degree of morphological plasticity. Shells of *B. longispina,* when small and have the elongate mucro are easily identifiable, but small shells with intermediate mucros are more obscure. In the present study the majority of remains can be attributable to either *B. coregoni* or *B. longispina* based on the above remarks. *B. longirostris* was rarely found, and never in great numbers. Where it was impossible to distinguish between *B. coregoni* and *B. longispina,* these remains were assigned to *Bosmina* spp.

There is further confusion regarding the viability of *B. coregoni* and *B. longispina* species. Fryer (1993) states that *B. coregoni* has been incorrectly used and that species of the *B. coregoni* group should be known as *B. longispina.* This view is based on comparison with material from Scandinavia and by Dr. U. Lieder (see also Lieder 1983a; Lieder 1983b). A recent paper by Hellsten and Sundberg (2000) casts doubt on this assertion however. Using Random Amplified Polymorphic DNA to test genetic divergence between two populations of *Bosmina* species in Lake Östersjön, Sweden, the authors (Hellsten and Sundberg 2000) present evidence that supports the specific status of *B. coregoni* and *B. longispina* previously based on the morphological characteristics. Furthermore, the *Bosmina* group as a whole has undergone numerous revisions and changes in classification and the situation is further confused by differences between European and North America

examples of the same species (e.g. Goulden and Frey 1963; Deevey and Deevey 1971; Korinek 1971; Lieder 1983a; Lieder 1983b; Lacroix 1989; DeMelo and Hebert 1994a; DeMelo and Hebert 1994b; DeMelo and Hebert 1994c; Kotov 1996; Korinek, Sacherova, and Havel 1997).

Whilst confusion regarding the classification of these species remains it is difficult to adopt a strict taxonomic procedure for dealing with the identification of remains of *B. coregoni* and *B. longispina.* For the purposes of this thesis, *B. coregoni* and *B. longispina* have, wherever possible, been identified on the basis of the morphology of the remains in question. Where there is ambiguity as to which specific species an exuviae be assigned to, the nominate *Bosmina* spp. has been used.

Fryer (1993) remarks that *B. longispina* (including *B. coregoni*) and *B. longirostris* have markedly different environmental requirement. *B. longirostris* is found in enriched waters, *B. longispina* (including *B. coregoni*), on the other hand, is found predominantly in oligotrophic waters and is tolerant of very acid conditions. Given this distinction it may prove desirable in the future to merge the counts of *B. coregoni* and *B. longispina* if no further ecological information is gained from separating them. This will require further statistical analyses to describe the responses of the two species to the measured environmental gradients.

2.3.2 The large *Alona* species

Two large species of the genus *Alona* are routinely found in lake sediments; *Alona affinis* and *Alona quadrangularis.* They are closely related but can be distinguished from differences in headshield morphology, shell characteristics and properties of the postabdomen.

2.3.2.1 Alona affinis *(LEYDIG, 1860)*

The headshield of *A. affinis* is pointed behind and bears two large median pores surrounded and connected by a thickening of the chitin in this area. The two lateral pores are located approximately one inter-pore distance from the median pores, generally positioned opposite the lower of the two median pores. The lateral pores are likewise embossed in an area of thicker chitin, and in some specimens there is the appearance of the lateral pores being joined to the lower median pore by some form of chitinous

thickening. The shell of *A. affinis* is large and nondescript. Where the surface markings are clear fine longitudinal striations are visible. The postabdomen of *A. affinis* generally has less than 14 teeth, which end abruptly at the anal groove rather than decreasing steadily to the anal groove and continuing as groups of small setae as they do in *A. quadrangularis*. The remains identified in this study do not differ from the published descriptions (e.g. Frey 1959; Frey 1962b; Flößner 1972).

2.3.2.2 Alona quadrangularis *(O.F. MÜLLER, 1785)*

This species is similar in appearance to that of *A. affinis*. Indeed, in Frey's (1959) description of the taxonomic and phylogenetic significance of the headpores of the Chydoridae he remarks that some workers, e.g. Keilhack (1909) and Weigold (1910) (both cited in Frey 1959) have suggested that *A. affinis* is a subspecies of *A. quadrangularis*.

The headshield of *A. quadrangularis*, like *A. affinis*, pointed behind. In *A. quadrangularis* however, the posterior of the headshield is smaller and decreases to less of a point than that of *A. affinis*, which has a larger, more sharply-pointed crest behind. The other major difference between the two species is in the arrangement of the headpores. *A. quadrangularis* has three median headpores connected in a narrow channel. These are positioned similarly to those of *A. affinis*, though tend to be located slightly further from the posterior margin of the headshield. The lateral pores are located two inter-pore distances from the central median headpore.

The shells of *A. quadrangularis* can have either coarsely spaced longitudinal striae or no markings at all. There may also be irregular reticulae in the anterior-ventral area of the shell. The postabdomen generally has greater than 15 teeth, with decrease in size and continue past the anal groove as groups of small setae.

A. quadrangularis was found in many of the samples in the training set and the remains identified did not differ from the description above.

2.3.3 Small *Alona* species

It is difficult to distinguish between the small species of *Alona* on the basis of the shells alone. Often they are devoid of characteristic markings that might be used to identify species affiliations. Also the headshields of some species, namely *Alona rectangula* (Sars, 1861) and *Alona guttata* (Sars, 1862) cannot be easily distinguished from one another. A further frustration in this area is the formation of tubercules on the shells and headshields of some species which do not normally have them. In the samples discussed here headshields of *Alona rustica* (Scott, 1895) were found in tuberculate form as were shells of *A. rectangula* (which may be given the varietal name *A. rectangula* var. *pulchra* (Hellich, 1874) (Frey, 1962)). Where it was impossible to determine specific species affiliation, either through masking by detritus or the presence of tuberculate forms, the nominate Small *Alona* spp. was assigned.

2.3.3.1 Alona rectangula *(SARS, 1861)*

The headshield of *A. rectangula* has a blunt rostrum and is broadly rounded in the posterior margin. There are three median headpores located in the midline of the headshield, with the lateral pores being located between one and two inter-pore distances from the central median pore. The shell of *A. rectangula* bears the characteristic parallel longitudinal striae, though these are not always visible in every specimen.

2.3.3.2 Alona guttata *(SARS, 1862),* Alona guttata *var.* tuberculata

The headshields of *A. guttata* are not easily distinguishable from those of *A. rectangula.* They have the same number of median headpores positioned as in *A. rectangula,* as are the lateral headpores. *A. guttata* might have a less rounded posterior margin than *A. rectangula,* as well as being slightly smaller with a more rounded rostrum (Frey, 1959). It is unlikely that this is sufficient to distinguish between the two species (Frey, 1959).

Shells have a rounded posterior-ventral margin and sometimes irregular patterns of tuberculae may be present. Where the tuberculae are present the varietal name *Alona guttata* var. *tuberculata* Kurz 1874 has been used. The tuberculae are also present in the headshields of *A. guttata* var. *tuberculata,* which in all other respects resembles the headshields of *A. guttata* and *A. rectangula.*

2.3.3.3 Alona costata *(SARS, 1862) and* Alona rustica *(SCOTT, 1890-99)*

These two species are very closely related morphologically (Frey 1964). The main feature of note are the characteristic transversely elongated headpores that are unique to these species (and *Alona bicolor sp. nov.* which is a species known from North America (Frey 1964)).

Both species have three median head pores. Those of *A. costata* are generally small and surrounded by a chitinous thickening that joins and surrounds the three pores. The middle pore is closer to the posterior pore than the anterior one. The lateral pores are transversely elongated slits, which are as long as, if not longer, than the distance between the middle and anterior pores. Each of these pores leads to a sac-like structure that has a slender filament extending from the apical end. This filament is rarely seen in sub-fossil material. The postabdomen of *A. costata* generally has 9 or 10 major teeth that do not extend past the post-anal angle.

The headshields of *Alona rustica* bear the elongated lateral pores as in *A. costata*, though in *A. rustica* they are smaller and the sac-like structures less developed. The headpores, of which there are three situated on the midline of the headshield, are similar to those of *A. costata*: three small pores surrounded and connected by a chitinous thickening. Most of the specimens observed in the samples under discussion in this thesis had headpores that resembled three single pores connected by fine channels as in Frey's figure 15 (Frey 1964 §15, page 163). The smaller sac-like structure of *A. rustica* was not seen in the sedimentary remains. When this is the case, the elongated headpores resemble a pair of parentheses facing each other, rotated through ninety degrees.

The postabdomen of *A. rustica* has prominent pre- and post-anal angles, being widest at the post-anal position. The postabdomen tapers to the distal end that is rounded and protrudes past the basal section of the post-abdominal claw. It has 9 to 11 teeth of which seven or eight are found between the distal end and the post-anal angle (Frey 1964). A feature that characterises the postabdomen of *A. rustica* is the fine armature that continues after the last set of teeth as groups of setae.

Of the two species, only *A. rustica* was found in any great numbers. Postabdomens of *A. costata* were found but these may have lost the armature of setae that extends past the final set of teeth.

2.3.3.4 Alona intermedia *(SARS, 1862)*

Only the headshields of this species were identified from the sediment. If the shells were present then they were unidentifiable as such and will have been placed in the nominate group of Small *Alona* spp.

The headshields of *A. intermedia* are distinct. They have two large round median headpores located on the midline approximately one inter-pore distance from the posterior margin of the headshield. The median pores are again connected by a rather distinctive channel. The lateral headpores are located approximately one inter-pore distance from the median pores, about level with the posterior edge of the lower-most median pore.

2.3.4 *Acroperus elongatus* (SARS, 1862)

The shells of *Acroperus elongatus* have parallel, coarse, diagonal striae, with more-closely spaced longitudinal striae found between them. The posterior-ventral margin of the shell usually has one coarse tooth, but there may be as many as two or as few as no teeth exhibited in some specimens. The headshields have blunted rostrums and are covered in closely spaced longitudinal striae that curve around the headpore arrangement at the posterior margin. There are three headpores, again connected as in *C. rectirostris* and *A. harpae*, and the lateral pores are positioned far from the median pore arrangement, near to the margins of the headshield.

The postabdomen of *A. elongatus* is superficially similar to that of *A. harpae*. That of *A. elongatus* however, is longer, has more teeth and those teeth extend around the distal margin of the postabdomen, which lacks the sharp indentation immediately prior to the point of attachment for the postabdominal claw.

2.3.5 *Acroperus harpae* (BAIRD, 1835)

The general shape of the headshield of *A. harpae* resembles that of *C. rectirostris*. In profile however, the headshield of *A. harpae* is more curved and the posterior margin proceeds to a point as opposed to the rounded posterior margin in *C. rectirostris*. The three headpores are connected and are found in a small channel near the posterior margin. The shells have coarsely spaced diagonal striae that become curved in the anterior-ventral area of the shell and point towards the dorsal margin. Three small teeth are found at the posterior-ventral margin of the shell, though some specimens found in these sediments lacked this feature. The postabdomen has fairly parallel edges and is shorter than that of *C. rectirostris*, with approximately 15 teeth that extend past the anal groove.

2.3.6 *Alonella excisa* (FISCHER, 1854)

The shape of the headshields of *A. excisa* is similar to that of *A. nana* though *A. excisa* is generally larger. The entire headshield of *A. excisa* is covered in fine longitudinal striations and the rostrum attenuated. The major headpores are located in an elongated oval located on the midline of the headshield. The minor pores are located equidistant between the two major pores.

The shell of *A. excisa* is covered in a pattern of diamond reticulae. The reticulae contain fine longitudinal striae visible at high magnifications. A notch or tooth is also found on the posterior-ventral angle.

2.3.7 *Alonella exigua* (LILLJEBORG, 1853)

Alonella exigua has the distinctive diamond shaped headshield that most of the species in the genus *Alonella* have. *A. exigua* has a more rounded rostrum and posterior edge and the headshield is reticulated in the central area extending to the posterior margin. The two small major pores are located in an elongated oval area surrounded by the reticulate patterning on the headshield. The two minor pores are located on the midline approximately equidistant from the major pores.

The shell of *A. exigua* is also strongly reticulated and a granular patterning inside the reticulae is visible at high magnifications. A prominent tooth is located at the posterior-ventral corner of the shell. *A. exigua* was rarely found in the sediment samples analysed, with *A. nana* and *Alonella excisa* being the dominant forms of *Alonella* in the sediments.

2.3.8 *Alonella nana* (BAIRD, 1850)

This is a small and very distinct species well represented in the sediment samples analysed for this thesis. The headshield is diamond-like in shape with a sharply pointed rostrum. The prominent longitudinal striae occur in pairs and sweep around the headpore arrangement where they join in the anterior position. There are two major headpores located about one inter-pore distance from the posterior edge. The minor pores are small and are located on the midline, equidistant from the two major pores.

The shells of *A. nana* have distinct parallel striae that sweep round and down towards the ventral margin in the anterior-ventral area of the shell. There is a prominent tooth on the posterior-ventral corner of the shell.

2.3.9 *Camptocercus rectirostris* (SCHOEDLER, 1862)

The shells and headshields of this species were uncommon in the samples analysed. The shell is generally oblong or rectangular in shape and has arched coarsely spaced striae. Between these striae are fine scratch markings. A number of teeth are found on the posterior ventral margin of the shell. The dorsal margin overhangs to accommodate the headshield, which is roundly truncated behind. The three headpores are found near to the posterior margin of the headshield. These median pores are connected. The headshields also exhibit the fine scratch marks that characterise the shells of this species. The postabdomens are long and slender, bearing *ca.* 15 teeth.

2.3.10 *Chydorus piger* (SARS, 1862)

The headshield of *Chydorus piger* has a short, rounded rostrum anterior to well-developed fornices. The width of the headshield expands after these fornices and the posterior end is

broadly rounded. There are two major pores located on the midline approximately central between the fornices and the posterior edge. Two minor pores are located between the two major pores. The shells of *C. piger* are oval and carry a distinctive, triangular shell flap in the anterior area. Sometimes radial striae are present on the shells, as are the strong ventral setae, which preserve well in sediments and are readily observable. Both the headshields and shells of *C. piger* were found in tuberculate form in some of the sediments analysed.

2.3.11 *Chydorus sphaericus* (O. F. MÜLLER, 1785)

Headshields of *Chydorus sphaericus* are distinguished from those of *C. piger* as they lack the well-developed fornices, have a large, more-tapered rostrum that is slightly notched at the tip, and a more broadly rounded posterior end. Head pores are located as in *C. piger*. Shells are again oval, though are generally reticulated and lack the triangular shell flap seen in *C. piger*. *Chydorus sphaericus* is a highly variable species but the features described are generally found in most specimens examined (Frey, 1959). The remains of other species were found in the sediment samples examined. These were generally rare and did not differ from previous descriptions of the taxa elsewhere in the literature. Therefore, these taxa will not be described here.

2.3.12 *Eurycercus lamellatus* (O. F. MÜLLER, 1785)

The remains of this species are very distinctive. The headshield has a single, large median headpore with the two lateral headpores closely adjacent. The postabdomen is gently curved and exhibits 100 or more saw-like teeth. The postabdominal claw is likewise large and gently curved, with two small teeth located at its base. The shell is large and has little or no distinguishing marks save for the posterior-dorsal margins of the carapace which are covered in tiny teeth.

The species *Eurycercus glacilis* Lilljeborg was looked for in the sediment, being distinguishable from *E. lamellatus* by the number of teeth (usually less than *ca.* 90) on the ventral margin of the postabdomen. All the postabdomens identified agreed with the assignment of *E. lamellatus* however.

2.3.13 *Graptoleberis testudinaria* (FISCHER, 1848)

Remains of *Graptoleberis testudinaria* were found in a number of sediment samples. *G. testudinaria* is easily recognisable by its distinctive headshield and shell. Both the headshield and shell have a distinct honeycomb patterning. The shell characteristically has two large teeth on the posterior ventral margin, though in a number of samples, specimens with 3, 4 and 6 teeth were recovered. Shells with this feature were in all other respects identical to *G. testudinaria.* This morphological plasticity has been noted in the literature, but the taxonomic significance or causal mechanism for this variation remains unclear.

The headshield of *G. testudinaria* has broadly rounded rostrum. The headshield narrows to a point dorsally. Three large median headpores are visible in the narrowing of the headshield, with the lateral pore visible prominently because of tubular thickenings of the chitin in these areas. The lateral pores were often obscured by detritus in the samples counted. Frey (1959) also describes the occurrence of small "accessory pores" which surround the median pores.

2.3.14 *Monospilus dispar* (SARS, 1862)

Monospilus dispar has both a characteristic headshield and a distinctive shell that are both easily recognisable in the sediment. The headshield has a single, irregular pore hole located near to the posterior margin. The headshield is broadly square in shape with an attenuated rostrum and a rounded tip.

The shell of *M. dispar* is rounded in the posterior and covered in faint reticulae. The faint setae along the ventral margin of the shell are sometimes visible in sub-fossil remains. The shell is not shed during moulting so shells are found nested. The postabdomen of *M. dispar* is small with a curved ventral edge. Approximately six teeth are present which curve round the distal end of the postabdomen and smaller setae are found between the termination of the teeth and post-anal angle. Small clusters of setae are also present on the sides of the postabdomen. The postabdominal claw is likewise small, with one spine found approximately one basal width from the base of the claw.

2.4 Numerical Methods

A range of numerical methods has been employed in analysing the work presented in this thesis. The techniques range from simple descriptive statistics and calculation of dissimilarity measures to the more complex ordination and canonical ordination techniques. Advanced permutation tests and forward selection procedures have been used to describe the statistically significant relationships in the various data.

A variety of software packages has been used to run the analyses. CANOCO for Windows (CANOCO version 4.5) was used for the ordination methods, with CANODRAW and CANOPOST being used to produce the ordination biplots presented here. MINITAB version 13.1 was used for the descriptive statistics, the plotting of data and for standardising data prior to analysis. R (Ihaka and Gentleman 1996) version 1.5.1 was used throughout the thesis in a data analysis and presentation role. Cluster analysis, generalised linear models, linear discriminant analysis and dissimilarity calculations were all performed using R. Some of the procedures for these analyses were taken from Venables and Ripley (1999).

2.4.1 Ordination Methods

Ordination methods are a range of techniques which order sites along axes according to their species composition. The term 'ordination' was first used by Goodall (1954, cited in Jongman, ter Braak, and Van Tongeren 1995) and comes from the German 'ordnung' which was first mentioned by Ramensky (1930, cited in Jongman *et al.* 1995) where he describes this type of approach.

A variety of ordination methods exists but the four main methods have so far been employed in analysing the diatom, Cladocera, and physico-chemical variable datasets. These are Principal Components Analysis, its canonical form Redundancy Analysis, and Correspondence Analysis and its canonical form Canonical Correspondence Analysis. The methods are described in more detail below.

2.4.1.1 Principal Components Analysis

Variation in the hydrochemical and catchment characteristic datasets was analysed using principal components analysis (PCA). PCA is an indirect ordination method that attempts to find the underlying hypothetical variables in a multivariate dataset. PCA can be seen to be an extension of fitting straight lines and planes by least squares regression.

If we wish to explain the response of a number of species to a given explanatory variable, e.g. pH, we could do so using a multiple regression of pH on each of the species. A least squares regression fitted to an individual species yields the residual sum of squares. This is a measure of how badly pH explains the distribution of the species. By fitting a least squares regression to each of the species in turn we produce a series of residual sum of squares for each species. Adding these values together yields the Total Residual Sum of Squares and is a measure of how badly pH explains the distribution of all the species.

If pH is the variable that best explains the distribution of the species from a set of explanatory variables then we might want to find the hypothetical variable that best explains the distribution of the species. PCA does this by constructing the theoretical variable that minimises the Total Residual Sum of Squares after fitting least squares regression to the data. PCA then finds the theoretical variable that next best explains the species distribution with the constraint that it is uncorrelated with the first. PCA then extracts the next best variable with the constraint that it be uncorrelated with the first two variables and so on. In PCA these theoretical variables are known as principal components or PCA axes, and can be equated with the correlation coefficients or eigenvectors of a variance-covariance or correlation matrix (Legendre and Legendre 1998).

PCA allows inter-correlations between explanatory variables to be identified. Each principal component or ordination axis is described by an eigenvalue. The eigenvalue is a measure of the amount of variation in the data that can be explained by the principal component.

The results of principal components analysis can be displayed graphically using a biplot or joint plot. In a biplot, vectors or biplot arrows are used to represent the explanatory variables and sites or samples are represented by points. The vectors for each explanatory variable point in the direction of maximal variation and the length of the vectors reflect the proportion of the total variance explained by the explanatory variable. The longest vectors

are the explanatory variables that are most important in determining the differences between sites or samples. The angles between vectors reflect the degree of correlation between explanatory variables. Acute angles between vectors indicate positive correlations between explanatory variables and obtuse angles between vectors indicate negative correlations. Vectors that lie perpendicular to each other indicate explanatory variables that are uncorrelated.

For a principal component analyses based on a variance-covariance matrix the biplot constructed from the results of the PCA also has the desirable property of preserving the Euclidean distance between objects (Legendre and Legendre 1998). The Euclidean distance is a geometric measure of the degree of dissimilarity between objects. On a Euclidean distance biplot, objects that are similar with respect to the explanatory variables will be clustered together, where as objects that are dissimilar will be separated in the ordination space.

Principal components analysis assumes a linear response by objects to explanatory variables. Therefore, it is an ideal technique for identifying relationships between the distribution of lakes and hydrochemical variables. It is also a useful technique for interpreting species distributions with respect to sites where the gradients in the species data are short e.g. less than 2 standard deviations, where the species response can be assumed to be linear. PCA is not robust to data that contains many zero values, so using PCA to analyse sparse biological matrices is inappropriate and other ordination techniques should be used (Jongman *et al.* 1995; Legendre and Legendre 1998). For stronger gradients, other indirect ordination techniques exist such as Correspondence Analysis that assume a unimodal response to explanatory variables that more accurately reflect the observed response of species.

PCA was implemented using the computer program CANOCO for Windows (CANOCO version 4.5 (ter Braak and Smilauer 2002)) and the number of principal components retained for analysis was determined using the Broken Stick model (see Jackson 1993). PCA was used to analyse the physico-chemical data sets and to analyse the Cladocera data where short gradient lengths in the data suggested linear responses to the underlying environmental gradients, which is the response model implicit in PCA.

2.4.1.2 Correspondence Analysis

Correspondence analysis (CA) and its detrended version (Detrended Correspondence Analysis or DCA) were used to investigate the distribution of samples and species and to uncover the main floristic variation in the species data.

Where PCA is based on a linear response between objects and explanatory variables CA (and DCA) is based on a unimodal response model. Under the unimodal model explanatory variables vary with respect to objects in a 'bell-shaped' response curve. Solving the Maximum Likelihood (ML) answer to fitting unimodal response curves to many explanatory variables and many sites is computationally intensive, so an alternative method of weighted averaging has been developed to approximate the ML results (ter Braak 1987) and this forms the basis of CA (Hill 1973).

Species commonly show a unimodal response to environmental gradients. For example, a diatom species may prefer a particular lake-water pH and may not exist at all in lakes with water that is too acid or too alkaline. Each of the species is largely confined to a specific interval of pH values. As before with PCA, we can build up a model of how well pH explains the species data. By taking the average of the pH values of the sites where the species is present, the distribution of species along the pH gradient can be described. This average, called the species score, is an estimate of the optimum pH value for the species, though it is a far from ideal estimation. To measure how well pH explains the species data we can look at the spread of the species scores along the pH gradient, known as the dispersion. If the dispersion is large then pH separates the species curves and explains the variance in the species data well. If the dispersion is small, then pH explains the variance less well. To compare the explanatory power of other environmental variables, each must be standardised before analysis.

If pH is the environmental variable that best explains the dispersion of the species data we might want to consider if there was another environmental variable that could have been measured that could explain dispersion in the species data still better. Correspondence analysis extracts the theoretical variable that best explains the species data.

CA does this by choosing the best site scores, i.e. the values that maximise the dispersion of the species scores (Hill 1973). This theoretical variable is known as the first ordination axis

of CA, or the first CA axis. Second and further axes can be extracted from the data; however they have the constraint of not being correlated with the first CA axis. This is to ensure that additional information is expressed on these later axes. Only a few axes are required in the hope that they represent most of the variation in the species data.

CA is a robust ordination technique for analysing ecological data that usually have large numbers of species (explanatory variables) and many zero values (Hill 1973; Jongman *et al.* 1995; Legendre and Legendre 1998).

Biplots can again be used to display the results of the CA. They are interpreted differently from those constructed from a PCA. Both species and sites can be positioned on the same Joint Plot, or they can draw separately as Distance plots. The actual properties of the plots are dependent upon the type of scaling used in the CA algorithm. Joint Plots are constructed using Hill's scaling where the emphasis has been placed on the positions of the samples in the ordination. Samples are positioned at the centroid of the species that are present in that sample (ter Braak 1987; ter Braak and Prentice 1988; Jongman *et al.* 1995; ter Braak and Smilauer 2002). In a Joint Plot species are positioned with respect to the optimum value for the taxon, abundance then declines in concentric circles away from the location of the species on the plot. In a biplot, however, species are positioned with respect only to their approximate species values (ter Braak and Smilauer 2002). The ordination diagrams preserve the Chi-Squared distance amongst objects and amongst predictors. Where emphasis has been placed on the samples, inter-sample distances reflect the Chi-Squared distance. When emphasis is placed on the species, the relative positions of the species approximate the Chi-Squared distance.

Correspondence analysis encounters problems when there is one axis that fully explains the species data. In such cases a spurious second CA axis is extracted that explains as much information as the first. This produces an arch defect in the ordination, which also leads to compression of points located at the ends of the gradient (Hill and Gauch, Jr. 1980; Gauch, Jr. 1982). A method of removing the arch and compression effects was proposed by Hill & Gauch (1980), known as Detrended Correspondence Analysis (DCA). A number of methods exist for detrending a CA but the preferred method (as it removes the arch effect as well as the compression of points at the end of axis 1) is detrending by segments. Effectively, the first axis is divided into a given number of segments, and then the mean

value for each segment is subtracted from each point in the segment (Jongman *et al.* 1995). This drops points down, reducing the strength of axis 2.

Non-linear rescaling is used to remove the distortion of points at the ends of axis 1. Non-linear rescaling is based on the assumption of a species-packing model and the fact that at the edges of a gradient, species curves are often truncated leading to correspondingly smaller within-site variances. Segments that have a small within-site variance are expanded, and those with a large within-site variance are contracted (Hill and Gauch, Jr. 1980). After standardising within-site variance, the site scores are then calculated as weighted averages of the species scores and are standardised so that within-site variance is 1 (Hill and Gauch, Jr. 1980). Now the length of the axis is the range of site scores in 'standard deviation' units. This has the benefit of allowing the length of gradients to be determined so that the suitability of unimodal or linear models can be assessed.

DCA has been used in this thesis to estimate gradient lengths (indications of species turnover along hypothetical environmental gradients) in the species data sets so that the appropriate ordination technique could be applied (e.g. linear or unimodal-based methods). DCA has also been used to analyse the diatom data to remove an obvious arch in the CA results.

2.4.1.3 Canonical Correspondence Analysis

Canonical Correspondence Analysis (CCA) is a direct gradient analysis technique, which has been used to determine the main explanatory environmental variables that best describe the variation in the species data (ter Braak 1986).

Indirect methods, such as PCA, CA and DCA, are so called because the techniques describe the variance in objects in terms of explanatory variables. For example, CA can be used to describe the positions of samples in terms of species present in those samples. Interpreting the distribution of species and samples in ordination space is then done indirectly using supplementary environmental data.

Direct methods take the other approach. Supplementary information is used implicitly within the ordination technique to provide a direct interpretation of the distribution of

species and samples using environmental data. Canonical ordination techniques are designed to detect the patterns of variation in the species data that are 'best' explained by the observed environmental variables. Ordination diagrams now show both the main patterns of variance within the species data and the relationships between the species and the environmental variables. Canonical ordination techniques combine ordination and regression into one technique. CCA is a constrained form of CA where axes are constrained to be linear combinations of environmental variables (ter Braak 1986).

If we continue the example given above for correspondence analysis, pH is the environmental variable that best explains the variance in the species data. Now we wish to find the theoretical variable that explains the variance better still. This time, however, this theoretical variable should be constrained to be a linear combination of the measured environmental variables. This is because in indirect gradient analysis techniques, correlations between the species and the measured environmental variables might only show up strongly in later axes (Jongman *et al.* 1995). As the technique of CA attempts to reduce the data down to as few axes as necessary this information may be overlooked. Therefore, we should consider the environmental variables from the beginning.

Perhaps some combination of pH and water temperature might explain the species data better than pH alone. After standardising the variables it is discovered that 3 x pH and 2 x water temperature disperse the species more than pH alone. Therefore, we should consider all combinations of environmental variables as well as the individual environmental variables themselves. CCA selects the linear combination of environmental variables that maximises the dispersion of the species scores, thus forming the first CCA axis. Second and further axes of linear combinations of environmental variables can be extracted, with the constraint that they are uncorrelated with previous CCA axes. As many axes can be extracted as there are environmental variables.

It can be seen, therefore, that CCA is a constrained form of correspondence analysis (ter Braak 1986). The restrictions become less strict the more environmental variables are included in the analysis. If the number of environmental variables is equal to, or greater than, 1 minus the number of species then the restrictions become zero and CCA is simply CA. The same problems with the arch and edge compression effects that can affect CA can also affect CCA (Jongman *et al.* 1995). The methods of detrending by segments or

polynomials available for CA can be applied to CCA. But in CCA superfluous environmental variables can be deleted from the analysis in order to reduce the arch effect (Birks, *pers comm.*). This is more desirable than detrending.

The results of a canonical correspondence analysis can be plotted in joint plots and triplots. In the same figure the weighted averaged species scores, the weighted averages or linear combination scores of the sites, and biplot arrows or class centroids of environmental variables can all be displayed. The length of the biplot arrow shows the strength of the variable with the direction of the arrow showing the major direction of variation.

CCA is a very robust technique. The major assumption is that it uses a unimodal response model so should be used where strong gradients in the species data are encountered. Where as CA approximates the Maximum Likelihood solution of a Gaussian ordination, CCA approximates the Maximum Likelihood solution of a Gaussian ordination if the correspondence analysis axis is close to the linear combination of environmental variables (ter Braak 1986; ter Braak 1987; Jongman *et al.* 1995).

CCA (or DCCA) has been used in this thesis to relate the diatom species data to the measured physico-chemical data. CCA was used in these cases because DCA and DCCA both showed that the diatom species data had long gradient lengths, which indicates that the species are responding in a unimodal way to the environmental data.

2.4.1.4 Redundancy Analysis

Redundancy Analysis (RDA) was invented by Rao (1964, cited in Jongman *et al.* 1995) and is the canonical form of PCA. Where PCA selects the theoretical variable that minimises the total residual sum of squares after fitting straight lines to the species data, RDA considers the linear combination of the measured environmental variables that minimises the total residual sum of squares.

The site scores, calculated from a weighted summation of the species scores, are regressed on the environmental variables, and the fitted values of this regression are taken as the new site scores in the RDA algorithm. The sites score are constrained to be linear combinations

of the environmental variables so RDA is just a form of PCA with restricted site scores (Jongman *et al.* 1995).

Ordination diagrams generated from the results of RDA can be interpreted in the same way as biplots. The species and sites points on the biplot together approximately represent the abundance of species at each site. Species whose points lie near to a site point will be present in high abundances at that site. The species points and environmental arrows represent the covariance between the species and the environmental variables (Jongman *et al.* 1995). Species points located at the centre of the biplot are not well fitted and inferring information about species abundances and correlations from these points is likely to be inaccurate.

The cladoceran data and the joint diatom and Cladocera training set have been analysed using RDA because the data suggest that the species are responding in a linear manner to the underlying environmental gradients (from DCA and DCCA analysis of the training sets). RDA is appropriate in such cases as the method assumes a linear species response model as the basis of the calculations.

2.4.2 Analogue Matching

Analogue matching is a palaeoecological technique that allows one to compare numerically the similarity between any two sub-fossil assemblages. In practice, the technique has been used most widely in the field of palynology as a tool for the quantitative reconstruction of past climates from fossil pollen spectra.

Flower *et al.* (1997) used the technique to identify those samples from a modern training set of lakes that were most similar to pre-acidification samples from two acidified Scottish lochs from the Galloway region (see sections 1.5 and 1.6 above for further details). The work described in Chapter 5 of this thesis has developed the approach to include two sediment proxies in the matching procedure and has investigated ways of determining the critical values used to determine the level of similarity required for any two samples to be declared similar.

This section of the thesis will describe the data analysis procedure and the techniques used in achieving the analogue matching results discussed in detail in Chapter 5.

A series of three modern training sets were produced; a 163-lake diatom training set, an 83-lake cladoceran training set and an 83-lake combined diatom and cladoceran training set. An 83-lake diatom training set was also created though this was not used explicitly in the analogue matching analysis presented in Chapter 5. This training set was used to examine the diatom species-environment response that would be included in the reduced training set for those lakes where sub-fossil cladoceran data were also available.

The training sets mentioned above are fully described in various chapters of this thesis. The 163-lake diatom training set, known as the UK Acid Waters Diatom Data Set (UKAWDDS) is described in Chapter 3. Chapter 4 discusses the 83-lake cladoceran training set, whilst the combined diatom and cladoceran training is discussed in Chapter 5, along with a description of the 83-lake diatom training set.

The species data generated from the analysis of sediment samples for diatom and cladoceran sub-fossil remains was compiled into data files using a Microsoft Access database application. This database was also used to process the associated physico-chemical data into formats suitable for data analysis. WinTran version 1.5 (Juggins, S., unpublished computer programme) was used to convert the species and physico-chemical data from Access SQL queries into cornel condensed format files read by CANOCO and ANALOG, amongst others. The export feature of Access was used to write comma-delimited text files for import into R version 1.5.1 (Ihaka and Gentleman 1996).

The diatom and cladoceran data sets were generated from the raw species abundance data and transformed into percentage data where the abundance of a particular taxon is re-expressed as the percentage contributed to the sample total. Rare taxa were defined as those taxa that contributed less than 2% to the sample total in a single sample in the training set. These taxa were deleted from the training set at the data processing stage. The percentages were not recalculated on the basis of the changed sample totals.

To generate the combined diatom and cladoceran data set the percentage abundance values in each of the diatom and cladoceran training sets were halved and the two training sets appended together. This resulted in a single species data file where the diatom taxa in a given sample contributed 50% to the sample total, as did the cladoceran taxa in the same sample. This process, in effect, gives equal weight to each of the species groups and the diatom and cladoceran abundance data contribute equal proportions to the sample totals. The procedure used in the thesis takes no account of the greater number of diatom taxa compared to cladoceran taxa in the training set. It should be noted however, that in removing the very rare taxa in the training set (see above), a larger number of diatom taxa were not considered for analysis compared to the number of Cladocera that were deleted for the same reason.

The computer program ANALOG (Birks & Line, unpublished computer programme) was used to run the analogue matching routine. ANALOG computes the dissimilarity between modern and fossil samples for a series of dissimilarity coefficients. ANALOG calculates the extreme percentiles of the distribution of dissimilarity values for each sample in the modern training set. These values can be used as a guide to the degree of closeness required for samples to be regarded as modern analogues.

The programme also performs two permutation procedures to test the results of the matching procedure against those one would achieve under a null model. The first permutation test permutes the columns of the data matrix (the species) and recalculates the row totals to percentages. The permutation test is run many times and after each permutation the dissimilarity between two randomly chosen samples is calculated. A permuted distribution of dissimilarity values is generated and the percentiles of this distribution are then used as an indication of the level of dissimilarity one might except for samples to be regarded as good modern analogues. For larger training sets with over fifty samples the distribution of observed dissimilarity values is a good guide to the percentiles, however, the distribution of dissimilarity values from the permuted data will provide a better approximation of the real percentiles for a smaller data set (Birks, H.J.B., *pers. comm.*).

R is a system for statistical computation and graphics (Ihaka and Gentleman 1996). The language of R is not unlike S as implemented in the commercial package S-Plus. R version 1.5.1 was used to calculate dissimilarity values and process the output into graphical form.

R was also used to generate the pseudo random numbers using to define critical values for the distribution of dissimilarity values. This is discussed in greater detail in Chapter 5.

2.4.3 Quantitative environmental reconstruction

The quantitative reconstruction of environmental variables from the sub-fossil remains of certain species groups (e.g. diatoms, chironomids, or pollen) has recently seen some of the greatest methodological advances in the field of palaeolimnology. The ability to provide accurate, quantitative predictions of past environmental conditions has been important in driving the research agenda within palaeolimnology and palaeoecology. Chemical targets for recovery from disturbance can be set using these reconstruction methods, and it is important to strike a balance between chemical and biological targets.

A variety of methods to perform environmental reconstruction has been developed. The majority of approaches fall into two general categories of methods, linear- or unimodal-based methods. This refers to the underlying taxon response model used in the various approaches and relates to the way species respond to the environmental variable being reconstructed be it in a linear or unimodal manner. These linear- or unimodal-based methods estimate the parameters of the regression model using a range of statistical methodologies (maximum likelihood, least squares, weighted averaging, partial least square, weighted averaging partial least squares etc.). In addition there are many of other methods that do not fall neatly into this taxon response model distinction, such as the modern analogue approach (MAT) and taxon response surfaces.

Chapter 5 makes particular use of MAT, and analogue matching *per se* forms the basis of MAT, and as such is used throughout this thesis.

2.4.3.1 Modern analogue technique

The modern analogue technique (MAT) is an inverse regression procedure where the fossil environment, x_0, is inferred directly from the calibration parameters, $\hat{\mathbf{U}}$, estimated from a regression of $\mathbf{X}$ on $\mathbf{Y}$ e.g.:

$$\mathbf{X} = \hat{\mathbf{U}}(\mathbf{Y})$$

and a calibration step:

$$\hat{\mathbf{X}}_0 = \hat{\mathbf{U}}(\mathbf{Y}_0)$$

Where $\mathbf{Y}$ is an n site by m taxa matrix of abundances, $\mathbf{X}$ is an n site by p variable matrix of environmental variables (p is usually taken to be 1), $\hat{\mathbf{U}}$ are the calibration parameters, $\hat{\mathbf{X}}_0$ is the n by 1 vector of inferred values, x_i, given the matrix of fossil abundances of y taxa, $\mathbf{Y}_0$.

MAT, or k-Nearest Neighbours (k-NN), as the method is also known, has been used in discriminant analysis problems to calibrate qualitative, rather than quantitative, variables. It has recently been used more often following a rise in the popularity of non-parametric smoothers (e.g. LOWESS and Generalised Additive Models (GAMs)) and k-NN can be seen as just one of the many ways of performing smooth regression (ter Braak 1995).

MAT is perhaps the least complex method of calibration currently available. Take as the estimate of the fossil value $\hat{x}_0$ the value of x_i of the particular sample in the modern training set that has the most similar species composition, y_i, to that of the fossil sample, y_0 (ter Braak 1995). The method may be extended to return the weighted average of the k most similar modern samples to y_0 as the inferred value $\hat{x}_0$. The weights are taken to be the inverse of the dissimilarities between y_i and y_0. An estimate of the standard error of the method can be made by taking the weighted sample variance among the k neighbours.

k-NN is one of a number of ways in which regression by way of smoothing can be performed. It is an inverse regression because it estimates the local mean of environmental variable x for the fossil species composition y. k-NN is a less sophisticated approach to smoothing than locally weighted smoothing, yet the two methods will return approximately the same results if the fossil sample is symmetrically surrounded in all dimensions of the y space by samples from the training set. If, however, the fossil sample lies close to the edge of the y space defined by the modern training set, the two methods will return different

results as *k*-NN will shrink towards the mean of x in the training set to a greater extent than locally weighted methods (ter Braak 1995). This bias in the uni-dimensional case becomes much more of a problem in multi-dimensional space because the perimeter of the y space is much greater.

The problems of high-dimensionality do not appear to effect *k*-NN in the same way that they do in SILR (segmented inverse linear regression) and locally weighted methods. In high dimensional spaces, points within that space will be sparsely distributed, with the result that, at least on some dimensions, the nearest neighbours will be located far apart. This is not the case in *k*-NN because the relevant dimensionality is p environmental variables, not m species, as it is in SILR and locally weighted methods. In addition, in *k*-NN the dimensionality just sums over species, the dimensions. In locally weighted and SILR methods, however, the local fits are based on at least m degrees of freedom, and are, therefore, more sensitive to the number of species, m (ter Braak 1995).

2.4.4 Cluster techniques

Sorting objects into groups is the first step towards developing conventions for classification. Inherent in the concept of classification is the belief that discontinuous groups occur in nature which is most often continuous in its appearance.

A large variety of clustering techniques has been developed. The various methods can be classified into two main types: Hierarchical clustering methods and non-hierarchical methods. Hierarchical methods can be agglomerative or divisive. These two types differ in that agglomerative clustering methods start with as many clusters as there are samples and then fuses the two samples that are the most 'similar' into a new cluster. The algorithm continues until only one cluster remains or until a finishing criterion is met. The term 'similar' can be described in many ways using dissimilarity coefficients and different linkage methods that control the way in which samples and clusters are fused together. Divisive methods on the other hand start with the entirety of the data and then split the data into two or more clusters. Subsequently, each cluster or sub-group is further divided until a finishing criterion is met.

Non-hierarchical methods attempt to find a predetermined number of groups of objects. All the groups are formed at the same level and thus there is no hierarchy. *k*-means clustering is an example of a non-hierarchical clustering technique, and has been used in this study to partition the environmental data for the lakes into clusters of different lake types. *k*-means is an iterative relocation method. Iterative relocation methods find an optimal solution to the problem of partitioning the samples into *k* groups, with *k* being known. The optimal solution is found by choosing appropriate staring points for the algorithm (e.g. cluster means from another hierarchical method, or at random based on a supplied number of cluster, *k*). The algorithm then moves samples between clusters so that the within-group variance is minimised and the between-group variance is maximised.

2.4.5 Modelling species responses to environmental variables

2.4.5.1 Generalised Linear Models

In a simple linear regression model we wish to estimate the values of α and ß for the following equation:

$$y = \alpha + \beta x + \varepsilon$$

where y is our response variable, x is the predictor variable, α and β are unknown but constant parameters (the systematic component), and ε is the error component that should be minimised. Traditionally, α and β are estimated using a least squares procedure which finds the values for the parameters α and β that minimise the sum of squared errors between the observed data and the predicted values.

In the least squares estimation procedure the error component, ε, is assumed to follow a normal, or Gaussian, distribution. Our data might not be normally distributed, however. They may be strongly skewed or kurtotic, or perhaps bounded in the upper or lower extremes or both (such as presence/absence or percentage data). Furthermore, when dealing with count data, as ecologists frequently do, the fitted values cannot be negative as a count of less than zero is clearly not sensible. As a result such data often violate the

assumptions of least squares estimation and as such are not suitable for analysis by such methods.

McCullagh and Nelder (1983) proposed an extension to the general linear model (of which linear regression by least squares estimation is just one form) called the generalised linear model or GLM. GLMs allow one to use any member of the exponential family of probability distributions as the error function in a model and account for the non-linear way in which the response variable often relates to the estimated model parameters.

To take account of these two changes we need to choose an error distribution that is appropriate for the data in hand, and to define two new functions that will take combinations of predictor variables and unknown parameters and then transform these combinations into the non-linear form of the response variables. The first function is the *linear predictor*, η, as the linear combination of the predictor variables and the unknown parameters β_i:

$$\eta = \sum_{j=1}^{m} x_{ij}\beta_i$$

The linear predictor can be used to predict y, the response variable, through the use of an intermediary, non-linear function known as the link function. The link function, $g(\ )$, allows the mapping of the linear predictor on to the scale of the response variable (McCullagh and Nelder 1983). The fitted values of the regression are calculated by applying the inverse of the link function to transform the linear predictor back on to the scale of the response variable using the following equation:

$$y = g^{-1}(\eta) + \varepsilon,$$

or,

$$g(y) = \eta + \varepsilon$$

where $g(\ )$ is the link function and η is the linear predictor (McCullagh and Nelder 1983).

So we have replaced the systematic part of the linear regression equation with a linear predictor and a link function. The linear predictor contains the best weights for the individual predictor variables, whilst the link function transforms these linear weights on to the non-linear response. A number of link functions are available for use with the various distributions and the choice of link is restricted to those that will provide a suitable transformation of the linear predictor on to the scale of the response variables for the error distribution used. Poisson, probit, log, logit and identity link functions are just some of the many types of link available.

Gaussian logit regression (GLR) is a special form of GLM and has been used in Chapter 4 to model the response curves of the Cladocera to the measured environmental data. GLR was used in preference to GLM with a Poisson error distribution because it was deemed more appropriate to analyses species presences and absences rather than species abundances (counts). The species abundance data generated for this thesis, instead of being discrete count data are percentage data and as such are closed, taking minimum and maximum values. The percentage data are different to raw abundances because taxon A might be, for example, present in two samples at different abundances, say 100 remains per gram dry weight of sediment and 1000 remains per gram dry weight of sediment. If the second sample contains a wide range of similarly abundant taxa that it is possible for taxon A to have similar percentage abundance in both samples, despite being numerically more abundant in the second GLR deals with binary data, which is ideal for analysing species presence absence data.

In GLR the response variable is the presences or absence of a particular species coded as 1 or 0 respectively. The predictor variable, x is one of the measured environmental parameters, a continuous variable. We now wish to model the presence/absence of a particular species as a function of x. The GLR model is defined by the following equation:

$$p_{ik} = c\exp\left[\frac{-0.5(x_i - u_k)^2}{t_k^2}\right]$$

where p_{ik} is the expected proportional abundance of taxon k in sample i, x is the measured environment in sample i, u_{ik} is the optima of taxon k in sample i, and t_k is the tolerance of taxon k (Jongman *et al.* 1995).

In GLM notation we can write the above equation as:

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 x + b_2 x^2 + \varepsilon$$

where $\log\left(\frac{p}{1-p}\right)$ is a logit transformation of the abundance, p.

Now the full power of the GLM can be used to model this relationship by selecting a binomial error function because the data are dichotomous, 0 or 1, and choosing a logit link function so that the raw species presence/absence data can be modelled as a function of the environment without out having to transform the species data or take prior account of the bimodal error distribution in the data:

$$p = g^{-1}\left(\sum_{j=1}^{m} b_0 + b_1 x + b_2 x^2\right) + \varepsilon$$

where $g(\ )$ is a logit link function and $\sum_{j=1}^{m} b_0 + b_1 x + b_2 x^2$ is the linear predictor, η.

Chapter 3: Diatom distributions in acid-sensitive, upland lakes in the UK

3.1 Introduction

The shift in emphasis, away from the identification of lake acidification and its causal mechanisms, towards processes of, and prospects for, recovery from acidification dictates that we must re-evaluate models of surface water acidification. Previous research (Battarbee and Renberg 1990; Birks *et al.* 1990a; Birks *et al.* 1990b) has focussed on developing biological models that were able to identify the extent of, and the damage caused by, surface water acidification. Whether these models are now suitable for answering question surrounding recovery from acidification remains to be seen. Furthermore, new data collection exercises (Allott and Rose 1994; Allott *et al.* 1995; Allott and Monteith 1999 and unpublished ECRC data)have yielded extra information that may be of use in the developing research area of recovery.

In northern Europe, the SWAP-palaeolimnology program (Battarbee and Renberg 1990) developed a diatom-pH training set from diatom surface-sediment counts and associated water chemistry for 167 samples from oligotrophic, acid-sensitive lakes across the UK, Sweden and Norway (Birks *et al.* 1990a; Birks *et al.* 1990b). A predictive model of pH, using the response of diatoms to pH, was constructed from this training set. This predictive model, known as a transfer function, has been used extensively to reconstruct pH changes in lakes from many areas of Europe. Other transfer functions exist, for example AL:PE (Cameron *et al.* 1999) and PIRLA (Charles and Whitehead 1986), which have been developed for use in a variety of circumstances or countries where using the SWAP transfer function was inappropriate (e.g. mountain lakes [AL:PE] and North America [PIRLA]). The most recently the European Diatom Database Initiative (EDDI) project is aims to develop transfer functions for pH by combining the various diatom-pH training

sets developed in Europe, using harmonised taxonomy and methods. Researchers may then select a subset of the EDDI samples that best represent the region in which the transfer function is to be applied, and in this way create local transfer functions for diatom-pH reconstructions.

These transfer functions have one main feature in common. They were all developed for the process of reconstructing pH changes from the diatom flora preserved in lake sediments. The impetus behind these transfer functions was a desire to identify whether lakes had acidified, and then to identify the causal mechanisms for any acidification. One direct result of this is that these training sets are biased towards acid lakes, with low ANC. Few of the lakes in the SWAP training set, for example, are above pH 6 or with calcium levels greater than 200 μ eq l^{-1}.

When we come to consider identifying recovery targets for acidified surface waters this property of the existing training sets becomes a problem. Those lakes likely to represent pristine, minimally impacted sites with respect to acid deposition are those which are likely to have a higher pH coupled with low calcium (e.g. <50 μ eq l^{-1}). This type of lake is extremely under-represented in training sets like SWAP, and this data inadequacy may limit their use with respect to identifying appropriate recovery targets.

A further problem with the SWAP training set is that it represents lakes from a wide range of environmental and geographical locations. When identifying appropriate reference conditions as recovery targets we need to consider the wider ecological context of those lakes. Whilst the diatom flora in a typical, acidified, oligotrophic lake in the UK may be very similar to that of a similar lake in Sweden or Norway, biogeographical factors are likely to show up in differences in species composition of organisms in higher trophic levels, e.g. Cladocera, chironomids and other invertebrates, and aquatic macrophytes. So, training sets that encompasses a wide biogeographical range may not be the most appropriate training sets for use in identifying appropriate reference conditions. A training set from a single region would be more appropriate for use in this situation.

As a direct result of these limitations a new diatom training set has been created for the UK. This training set is based on the UK samples from the SWAP training set together with the addition of newly sampled sites from minimally impacted areas in the North West of

Scotland. This training set, known as the UK Acid Waters Diatom Data Set (UKAWDDS) forms the basis of the work presented in this thesis. The UKAWDDS will be the dataset used in palaeolimnological research on recovery and the definition of recovery targets, and the development of the analogue matching approach focuses on a subset of samples from the complete UKAWDDS. As such it is appropriate to subject this training set to a rigorous analysis to understand the properties of the data.

This chapter describes the properties of the UKAWDDS and presents some preliminary reconstruction techniques using the diatom-pH relationship. A comparison of this training set to other published training sets is also made.

3.2 UK Acid Waters diatom data set

The UKAWDDS contains a series of data on 163 samples from 151 UK lakes. The lakes contained in the training set are typical of those found in acid-sensitive areas of the uplands of the UK. The lakes are taken from Scotland (99 lakes), Wales (47 lakes) and England (5 lakes). Twelve of the lakes in the training set are represented by two surface samples. Separate hydrochemistry samples correspond with these duplicate diatom samples. They were taken a number of years apart and can be regarded as independent samples for the purpose of this study. The samples are not vastly different in terms of diatom species composition or hydrochemistry (Stevenson *et al.* 1991).

3.2.1 Physico-chemical data

Twenty-four physico-chemical variables were measured across the 163-sample dataset. The physical variables were available for the 151 lakes in the data set and the hydrochemical variables were available for the 163 samples. The hydrochemical data mostly represent means of the sampled water chemistry for the period over which the diatom sample represented (pH is a geometric mean of the available data). In most cases, this was the mean of four quarterly water samples. A few samples in the hydrochemical data represent a single water chemistry sample. Standard methods adopted by the Critical Loads Advisory Group (CLAG) (now reconstituted as the National Advisory Group on Transboundary Air Pollution, NEGATP) on behalf of the Department of the Environment, Food and Rural

Affairs (DEFRA) were used for the single samples. These were taken during the autumn period of the year, which is the best period of the year for collecting representative hydrochemical samples (Kernan, pers. commun.).

Summary statistics for the hydrochemical and physical data are shown in Table 3 and Table 4 respectively. In these two tables, the trimmed mean is the mean of the central 90% of the observations, after ignoring the top and bottom 5% of the values. The trimmed mean is less influenced by outlier or extreme values than the standard measure of the mean. These statistics show that lakes in the training set have a wide range of pH values (4.38 – 7.17 pH units) with similarly large ranges in calcium (11.47 – 390.7 $\mu eq\ l^{-1}$), equivalent alkalinity (-20.0 – 415.53 $\mu eq\ l^{-1}$), and total aluminium (0.0 – 340 $\mu g\ l^{-1}$). Summary statistics for conductivity and other related variables (e.g. sodium, chloride, magnesium etc.) also show wide ranges in the values across the dataset suggesting a strong conductivity-sea salt gradient in the hydrochemical data, which is superimposed upon the pH gradient.

Table 4 shows that there are equally large ranges in the physical data across the training set. The majority of the lakes have little or no afforestation in the catchment (median = 0.00), with 75% of the lakes having 20% or less afforestation (upper quartile [Q3] = 20.00). The statistics show that there is an extremely large range in lake area and catchment area values, although this is the result of two extreme values at the high end of the range. The results of the analysis shown in Table 4 indicate that the data set contains a wide range of different lake types; shallow to deep, low to high altitude, small to large.

Figure 6 shows frequency histograms of the hydrochemical data for the 163 samples in the UKAWDDS. With the exception of pH and TOC, the hydrochemical variables are strongly skewed to the right, being dominated by low values, and follow approximately a log-normal or Poisson distribution. As such all variables except pH and TOC were $\log_{10}$ transformed prior to subsequent analyses to stabilise the variance, to down-weight the influence of the extreme values, to reflect the linear response of many organisms to the log of chemical variables, and to approximate better the normally distributed random errors which many statistical analyses assume.

Table 3: Summary statistics for the hydrochemical data from the 163-sample UK Acid Waters diatom data set. (N = number of samples, N Miss = number of samples with missing data, Trim Mean = trimmed mean, Std. Dev. = Standard deviation, SE Mean = Standard error of the mean, Min = minimum value, Max = maximum value, Q1 = 1st quartile, Q3 = 3rd quartile).

Variable	N	N miss	Mean	Median	Trim Mean	Std. Dev.	SE Mean	Min	Max	Q1	Q3
pH	163	0	5.748	5.717	5.7838	0.7202	0.0564	4.382	7.174	5.17	6.37
Conductivity (μ S cm^{-1})	163	0	55.07	44.70	50.66	35.85	2.81	13.6	251.0	33.05	61.0
Total organic carbon (mg l^{-1})	145	18	3.251	2.850	3.111	2.121	0.176	0.12	10.5	1.55	4.45
Calcium (μ eq l^{-1})	163	0	90.86	63.00	83.45	72.66	5.69	11.47	390.7	38.45	118.0
Magnesium (μ eq l^{-1})	163	0	76.21	58.90	70.74	52.37	4.10	0.0	282.0	43.0	92.0
Sodium (μ eq l^{-1})	163	0	269.1	210.4	234.6	227.0	17.8	67.0	1808.7	157.0	288.8
Potassium (μ eq l^{-1})	163	0	9.915	8.000	9.101	6.623	0.519	2.299	40.55	5.961	11.50
Sulphate (μ eq l^{-1})	163	0	85.3	72.60	81.43	46.87	3.67	21.00	264.32	54.67	105.0
Chloride (μ eq l^{-1})	163	0	298.7	216.6	256.1	282.0	22.1	45.4	2094.8	155.2	303.0
Nitrate (μ eq l^{-1})	123	40	12.96	3.06	8.80	25.29	2.28	0.0	151.0	0.80	12.67
Alkalinity (μ eq l^{-1})	125	38	65.97	34.97	54.14	86.17	7.71	-19.0	434.45	18.74	78.50
Equivalent alkalinity (μ eq l^{-1})	159	4	51.24	18.33	38.51	82.87	6.57	-20.0	415.53	2.90	60.52
Total aluminium (μ g l^{-1})	162	1	63.55	40.50	57.34	62.31	4.90	0.0	340.0	13.0	98.95
Monomeric aluminium (μ g l^{-1})	114	49	25.35	17.63	22.42	24.98	2.34	0.0	132.6	8.46	32.30
Labile aluminium (μ g l^{-1})	114	49	22.82	7.000	18.38	32.89	3.08	0.0	207.0	1.50	34.04

Table 4: Summary statistics for the physical data from the 163-sample UK Acid Waters Diatom Dataset. (N = number of samples, N Miss = number of samples with missing data, Trim Mean = trimmed mean, Std. Dev. = Standard deviation, SE Mean = Standard error of the mean, Min = minimum value, Max = maximum value, Q1 = 1st quartile, Q3 = 3rd quartile).

Variable	N	N miss	Mean	Median	Trim Mean	Std. Dev.	SE Mean	Min	Max	Q1	Q3
% Afforestation	163	0	15.94	0.00	12.23	29.01	2.27	0.00	100.00	0.00	20.00
Lake area to Depth ratio	148	15	2.278	1.060	1.881	2.826	0.232	0.150	16.330	0.593	2.980
Catchment to lake area ratio	107	56	18.62	10.00	13.68	34.57	3.34	1.50	320.00	5.32	18.29
Catchment Area (ha)	107	56	693	96	260	2828	271	2	26200	39	310
Lake Altitude (m)	162	1	284	255.0	271.0	184.4	14.5	10.0	1000	149.5	396.5
Lake Area (ha)	148	15	28.66	11.00	21.04	55.38	4.55	1.00	473.0	5.00	34.00
Maximum Altitude in Catchment (m)	99	64	567	554.0	558.5	302.9	30.4	47.0	1296.0	334.0	764.0
Maximum Lake Depth (m)	161	2	12.601	11.00	11.723	8.759	0.690	1.000	51.000	6.900	16.50
Net Catchment Relief (m)	98	65	266.6	218.5	247.6	228.1	23.0	4.0	935.0	83.5	391.0

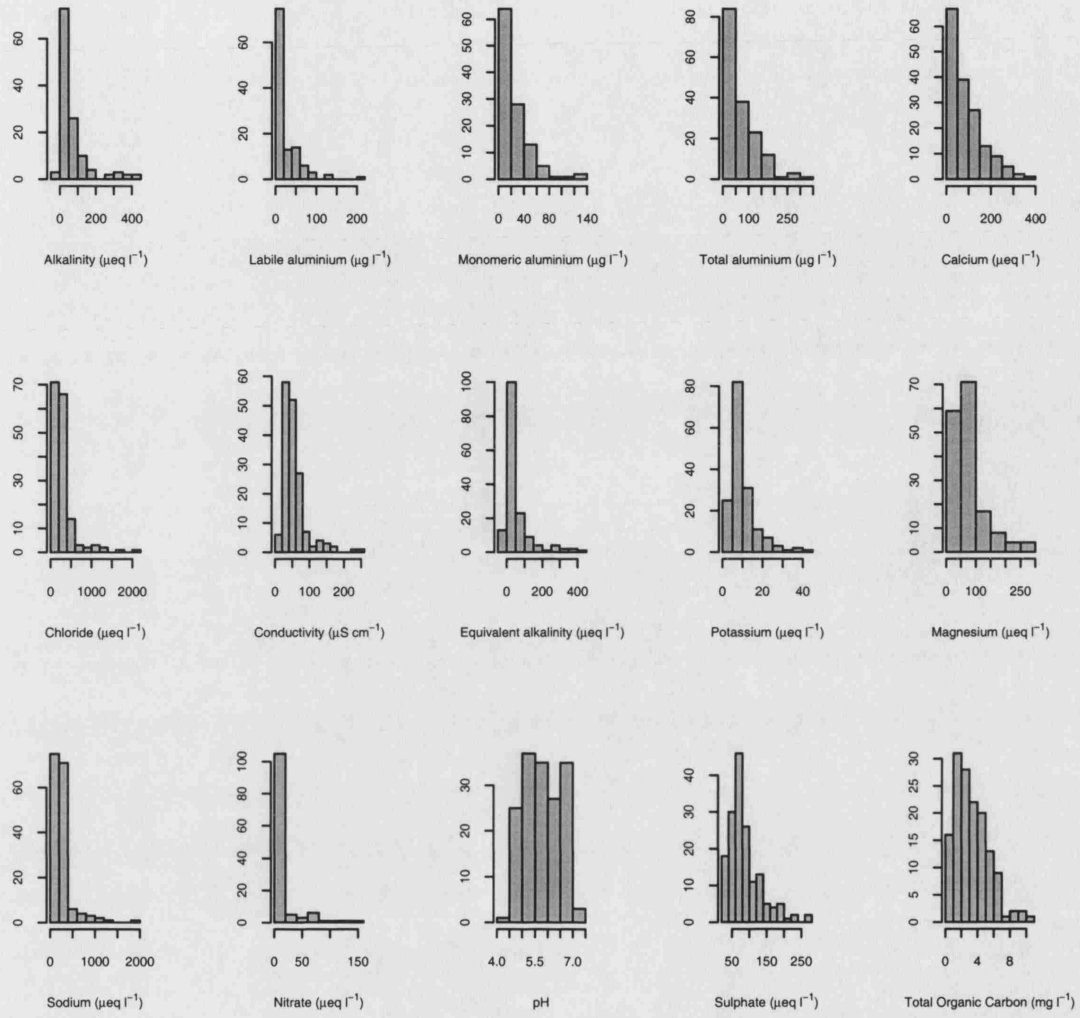


Figure 6: Frequency histograms of the hydrochemical data for the entire UK Acid Waters Diatom Data Set. Y-axis values are counts for each bin width. Bin width was selected automatically based on the number of bins, as determined by Scott's rule (Scott 1979).

The histograms show that the majority of the sites are acidic, with low modal calcium (<50 μ eq l^{-1}), conductivity (33 – 66 μ S cm^{-1}) and equivalent alkalinity (<50 μ eq l^{-1}) classes.

Figure 7 shows the frequency histograms for the 9 physical variables in the UKAWDDS. As with the hydrochemical variables, the histograms show that most of the variables are strongly skewed, varying widely from a normal distribution. Catchment area and lake area are strongly skewed, the result of a few extreme observations at the high end of the range.

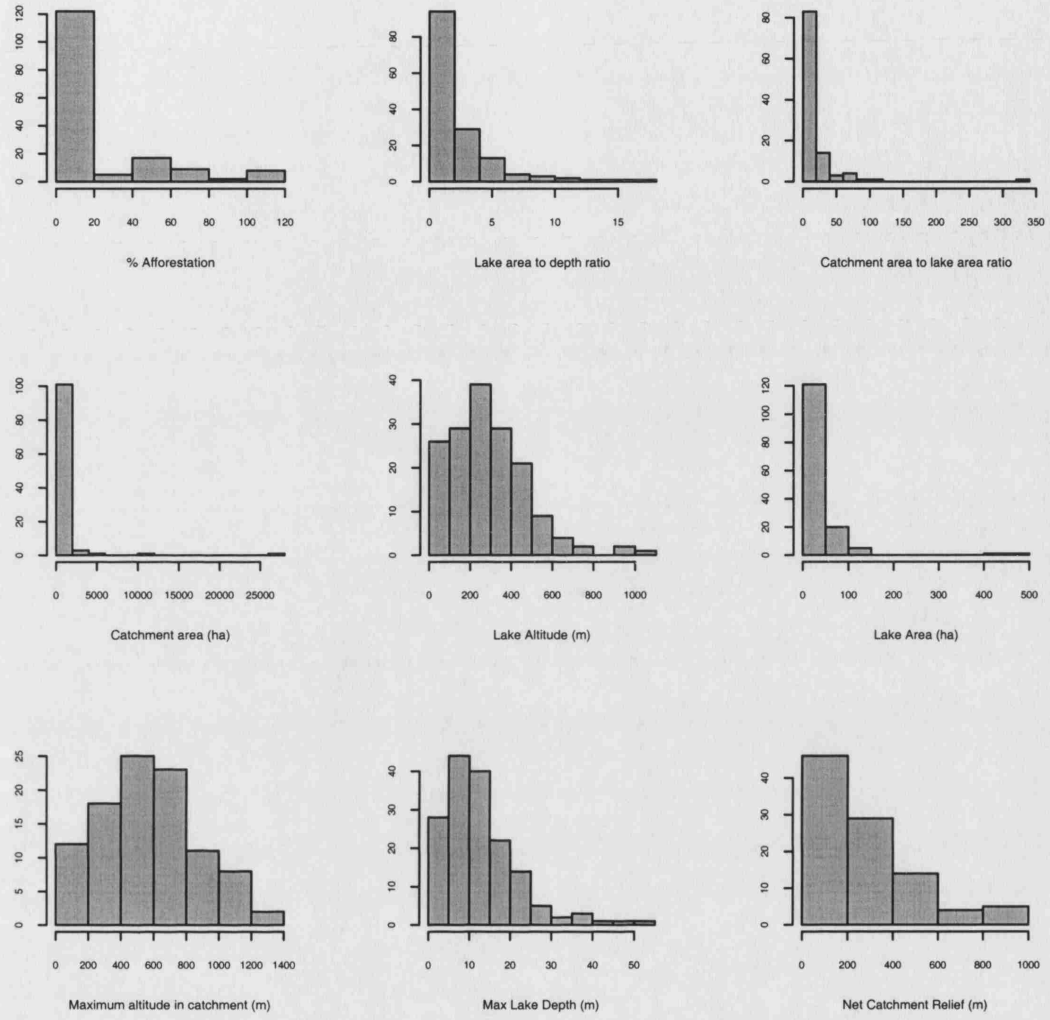


Figure 7: Frequency histograms of the physical data for the entire UK Acid Waters Diatom Data Set. Y-axis values are counts for each bin width (class). Bin width was selected automatically based on the number of bins, as determined by Scott's rule (Scott 1979).

Only maximum altitude in the catchment approximately follows a normal distribution. All the other physical variables were $\log_{10}$ transformed prior to subsequent statistical analyses. The histograms reflect the properties of the data illustrated in Table 4, namely that the majority of the lakes contain little or no afforestation, and are relatively small (mode <50 ha). The data cover a range of altitudes (0 – >1000m, mode = 300 – 350m) and maximum depths (0 – 51m, mode = 5 – 10m)

In order to investigate the physico-chemical data further a hierarchical agglomerative cluster analysis was performed on standardised (zero mean, unit variance) hydrochemical data using an unweighted group-average linkage method and the Euclidean distance

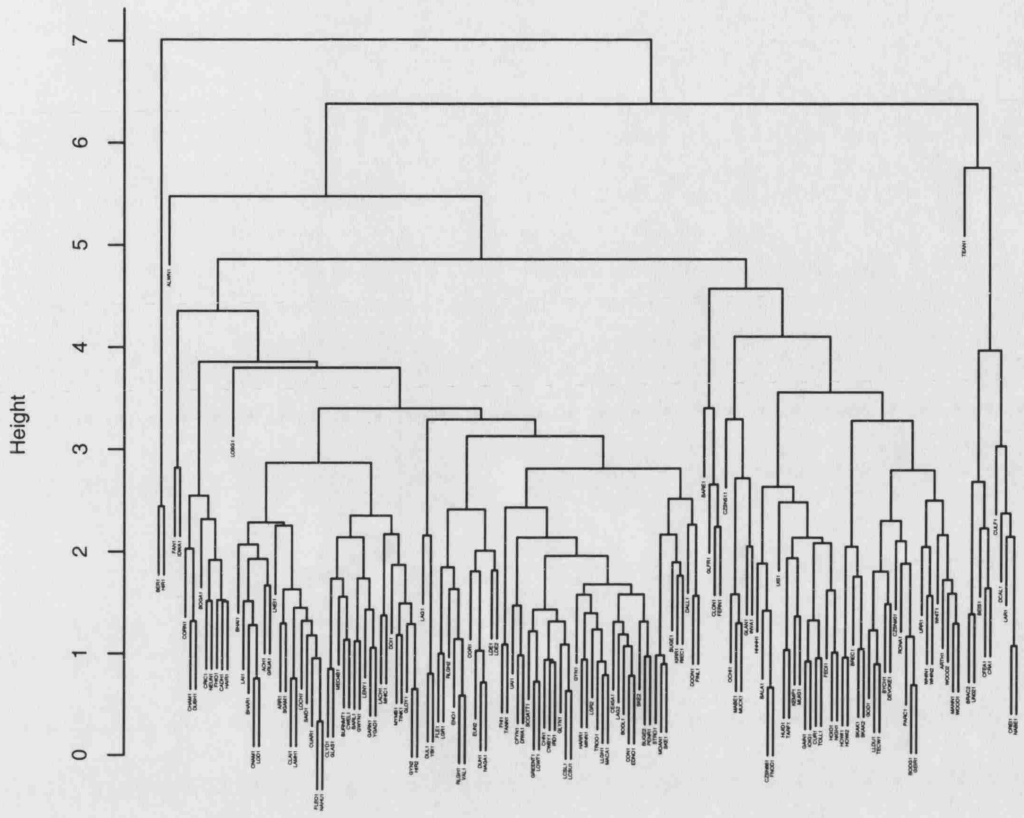


Figure 8: Dendrogram of a hierarchical cluster analysis of the 163 hydrochemical samples in the UK Acid Waters Diatom Data Set. See text for method.

measure. In an unweighted group-average cluster analysis, clusters are fused based on the average of all the distances between all the objects in cluster *k*, and sample *j*. As this is a hierarchical method, a dendrogram of the cluster analysis can be drawn. The dendrogram of this cluster analysis is shown in Figure 8.

The dendrogram shows the presence of 4 main clusters of samples, and a few remaining outlier lakes (Loch Teanga [TEAN1], Llyn Alwyn [ALWN1], Llyn Berwyn, [BER1], Llyn Hir [HIR1], Loch Laidon [LAI1], Llyn Tegid [BALA1], Loch Doilet [DOI1].

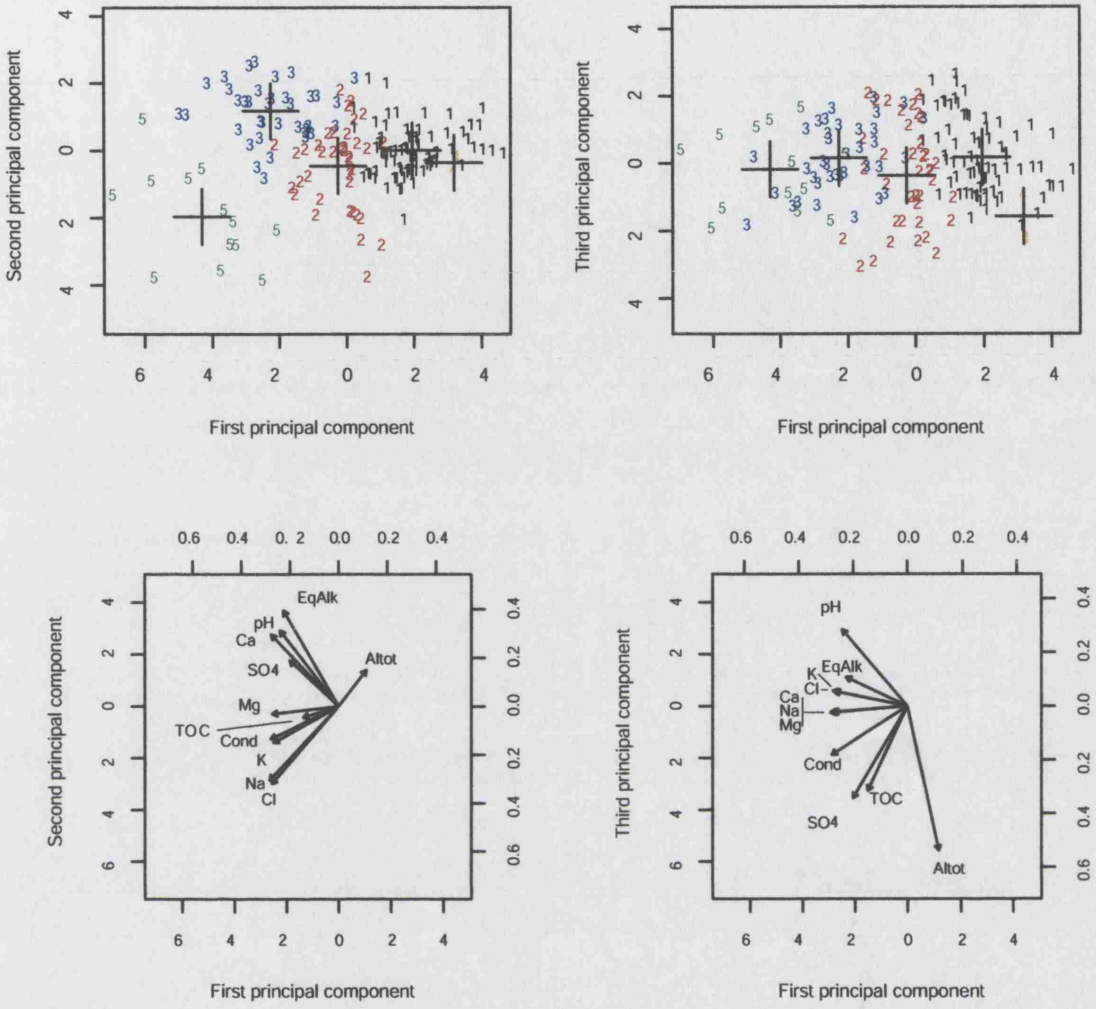


Figure 9: Diagram showing the results of the k-means cluster analysis of the hydrochemical data in the UKAWDDS. The results of the cluster analysis are projected into principal component space and the cluster centroid is plotted. Below each cluster plot, biplot arrows show the importance and direction influence of the hydrochemical variables along the plotted principal components. The first three axes of the principal components are significant under the broken stick distribution. The plot of the second against the third principal components is not shown because it demonstrated little in the way of cluster separation.

The unweighted group-average clustering method is hierarchical, and as such may not be the most optimal clustering of the samples in the physico-chemical data set. There is no suggestion that in nature a hierarchy of lakes exists. It is more likely that the lakes in the data set form a continuum of sites.

An alternative method of clustering is *k*-means clustering. *k*-means is a non-hierarchical method, and works by iteratively moving samples between clusters with the constraint that within-group variance is minimised and between-group variance is maximised.

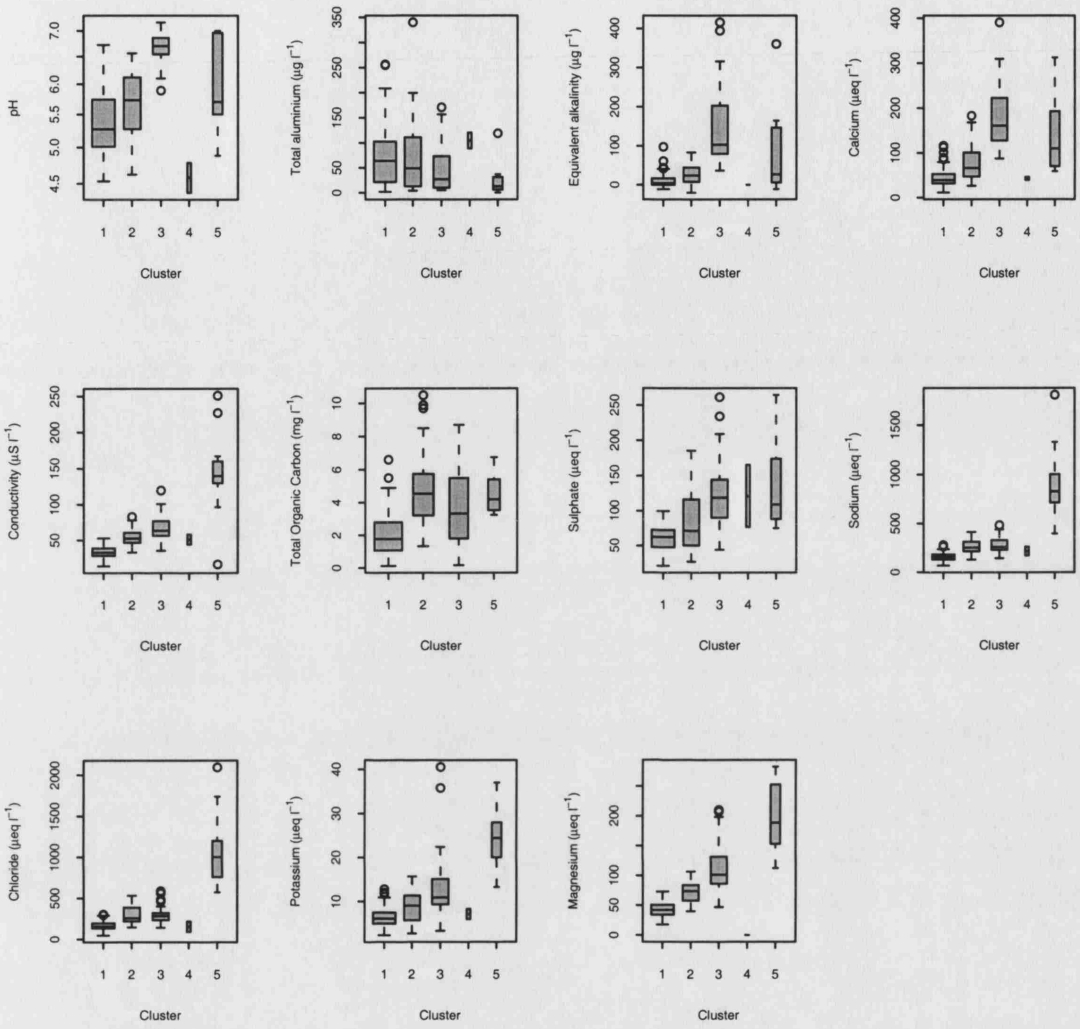


Figure 10: Boxplots of the hydrochemical variables plotted for each cluster identified by the k-means cluster analysis. The width of the boxes is proportional to the variance across the cluster of the variable.

The dendrogram was cut into five clusters: the four main groupings and the outlier lakes. The cluster means, or centroids, were then used as starting points for the *k*-means algorithm. As *k*-means is a non-hierarchical solution to the problem of clustering objects, a dendrogram of the results can not be produced. To visualise the results of the *k*-means clustering the groups were projected into the environmental space described by a principal components analysis of the physico-chemical data. This form of visualisation is shown in Figure 9.

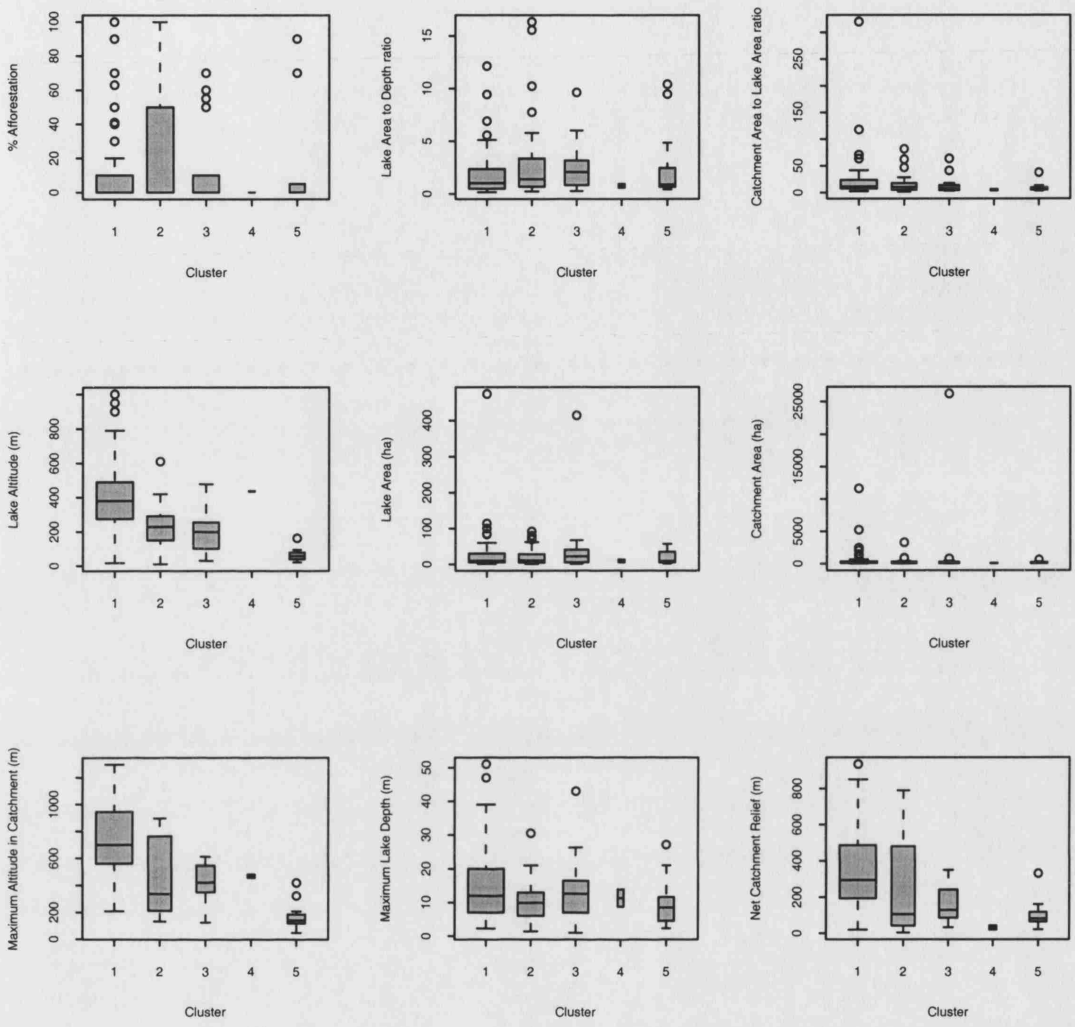


Figure 11: Boxplots of the physical variables for each of the clusters identified using k-means cluster analysis (163-samples). The shaded grey boxes are plotted so that the width of the box is proportional to the variance in that variable for each cluster.

Figure 9 shows projections of the 163-samples into principal component space. The projections shown are of the first and second principal components and the first and third principal components. Even though the first three axes of the principal components analysis were significant, as assessed by comparison against the broken stick model (Jackson 1993), the projection of the second and third principal components is not shown because there was little separation of the clusters in that environmental space.

The projection of the first and second principal components shows a good degree of cluster separation, with little overlap between clusters. However, the majority of samples in clusters 1, 2, 3, and 4 grad into one another. This is a reflection that, in nature, lakes occur

along a continuum rather than in hierarchical groups. The first principal component provides the best degree of cluster separation. The biplot arrows show this component to be a combination of the effects of conductivity (Na, Cl, K, Conductivity, Mg) and acidity (pH, Ca, and EqAlk). These variables are also strongly correlated with the second principal component. From this plot, and that of the first and third principal components, we can deduce that cluster five contains lakes that have higher than average conductivities, including high values for sodium and chloride. Cluster 4 is a small group, containing only two lakes, which is best described as being of below average conductivity, and lower than average pH. Clusters 1, 2, and 3 all have similar levels of lake water conductivity. However, the clusters are better separated along the acidity gradient. Cluster 3 contains lakes with high pH (low acidity lakes) with clusters 2 and 1 containing progressively more acid lakes.

The plot of the first and third principal components shows a similar pattern to that of the previous plot. The major factor determining cluster separation is the conductivity of the lakes in each cluster. The separation of the clusters in terms of the acidity of the lakes is also present, though aluminium has greater explanatory power along the third principal component then the other acidity related variables.

Figure 10 shows boxplots of the hydrochemical variables plotted for each cluster. These boxplots clearly show the hydrochemical differences between the clusters, emphasising the main trends shown in Figure 9 above. However, further properties of the clusters are now more apparent. Clusters 3 and 4 have higher calcium levels and equivalent alkalinity levels than the other groups. These two clusters contain the less-sensitive lakes in the dataset to acid deposition. The boxplot for conductivity again illustrates the strong separation f the clusters. Cluster 5 contains lakes with considerably greater conductivity levels than the other groups. Cluster 1 contains lakes with the lowest conductivities with clusters 2 and 3 containing lakes with progressively greater

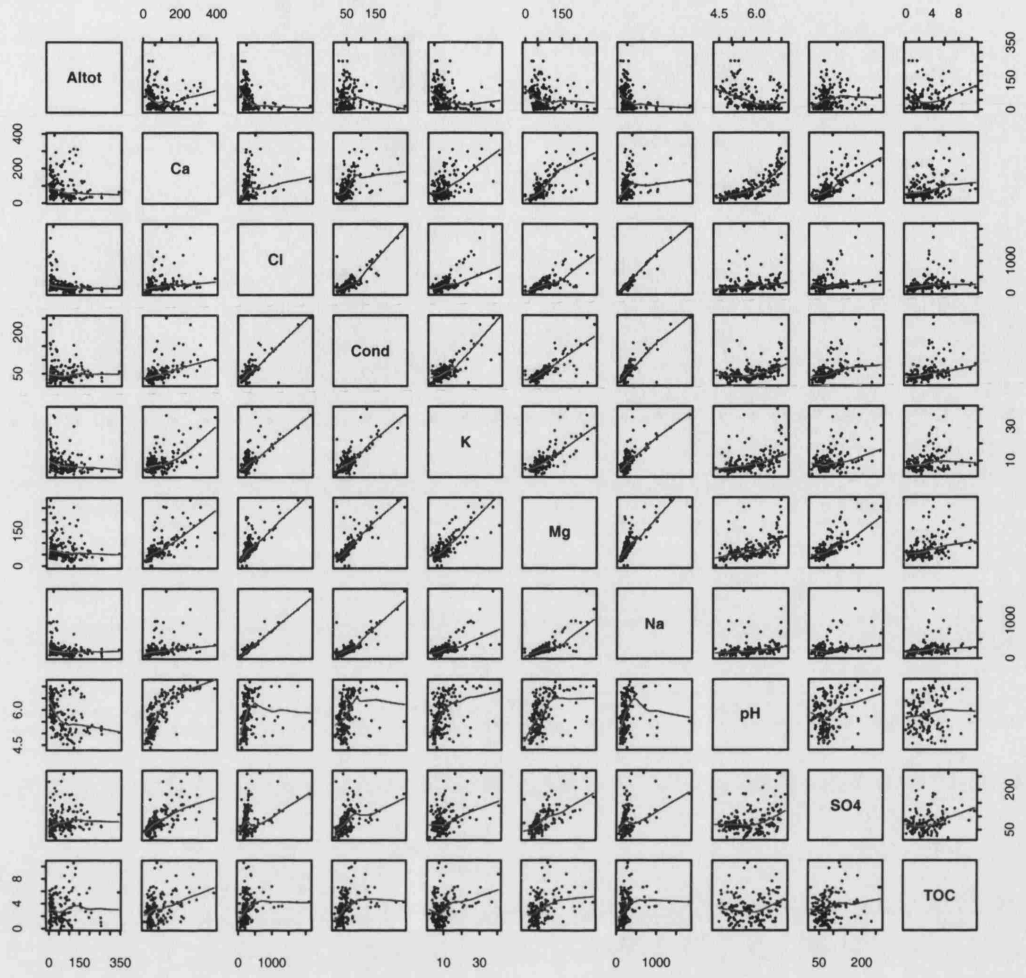


Figure 12: Scatterplot matrix of the 10 hydrochemical variables for which complete data was available. Overlaid on each plot is a LOWESS scatterplot smoother (span=0.66) illustrating the correlation between each pair of variables.

conductivity concentrations respectively. The boxplots for sodium, chloride, potassium and magnesium also show a similar trend to that shown in the conductivity plot.

Figure 11 shows boxplots of the nine physical variables plotted for each cluster identified in the *k*-means analysis. Again, it is difficult to separate the clusters using any one physical variable. The only variables that show differences between clusters are those related to altitude. Cluster 1 contains lakes found at the highest altitudes, with mean altitude declining in clusters 2 to 5. A similar, though less apparent, pattern is shown in the variables maximum altitude in catchment and net catchment relief.

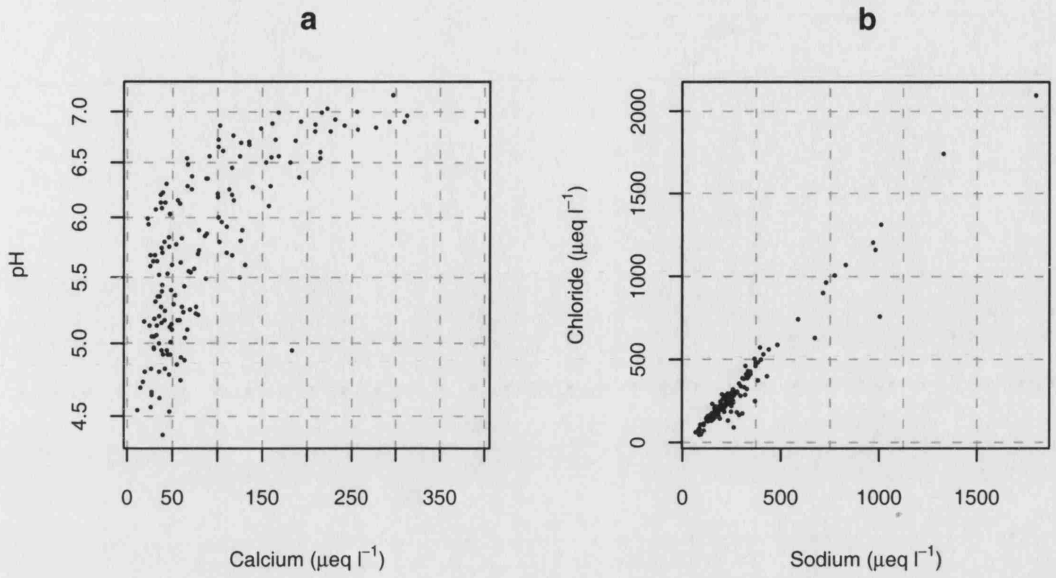


Figure 13: Scatterplots of calcium against pH (a) and sodium against chloride (b) for the 163 hydrochemical samples in the training set.

The results of the cluster analysis demonstrate that the lakes within the 163-samples in the dataset form a continuum of lakes from a broad range of physico-chemical conditions. Shows that lakes with different conductivity and acidity values occur across the range of physical parameters. For example, lake depth and area, and catchment area both appear to have little influence on the hydrochemistry of the lakes. Even the amount of afforestation in the catchment is broadly similar across the clusters. However, the more acid lakes do tend to be located in those lakes found at higher altitudes. This is to be expected as higher altitude lakes will have thin soils with lower acid buffering capacities, and this is born out somewhat in that the lakes with the lowest calcium and alkalinity levels are those lakes found at higher altitudes with greater relief in the catchment.

A scatterplot matrix of the hydrochemistry data are shown in Figure 12, showing the summary Scatterplot matrix, illustrating the correlations and interrelationships between the various hydrochemical parameters. Overlaid on each of the scatterplots is a LOWESS smoother (span=0.66) indicating the main trend in the data. As is to be expected, there are strong (*ca.* 1:1) correlations between conductivity of the water and the content

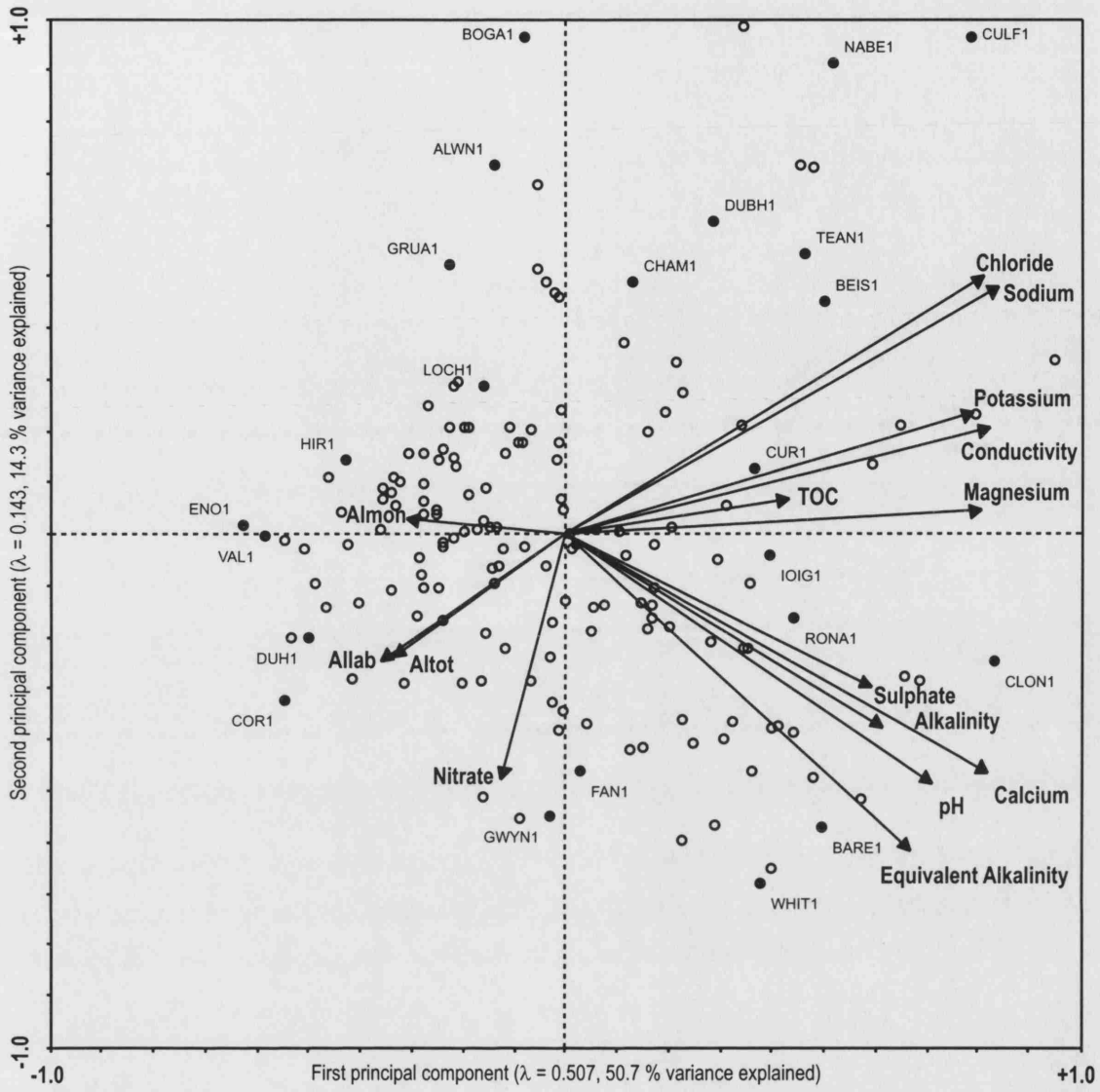


Figure 14: Correlation biplot showing the results of the principal components analysis of the 15-variable, 163-sample hydrochemical data set. The values in brackets are the eigenvalues for the respective principle components. To improve the clarity of the biplot only some of the samples are labelled (see appendix A for sample codes). Allab, Almon and Altot = labile, monomeric and total aluminium respectively. See text for interpretation.

of the major ions. There are also strong relationships between pH, total aluminium, Ca, TOC and sulphate.

One of the weaknesses of the SWAP dataset for use in identifying reference conditions for acidified lakes, was the limited coverage of lakes with high pH but low acid neutralising capacity in the training set. Figure 13a shows the relationship between pH and Ca content of the lake water for the 163-samples in the training set. It is clear from this figure that the training set contains a considerable amount of variation in pH at low (<50 $\mu eq\ l^{-1}$) calcium concentrations.

The training set contains a number of lakes with low calcium concentrations yet relatively high pH values. These lakes are likely to be minimally impacted sites, because they are sensitive to acid deposition yet have remained relatively unacidified. Figure 13b shows that across the data set sodium and chloride are in approximate marine proportions, though many values do deviate widely from this ratio, especially at the low end of the sodium scale, with many lakes containing a higher proportion of sodium to chloride.

Principal components analysis was used to further investigate the underlying environmental gradients present in the physico-chemical data sets. The hydrochemical and physical sets of variables have been analysed separately to determine the major environmental gradients for the respective sets of variables. A joint PCA of all the physico-chemical was also performed to investigate the relative importance of the hydrochemical and physical data across the data set.

The PCA of the hydrochemical data indicates the presence of two strong gradients in the data (Figure 14, Table 5). 86% of the variation in the hydrochemical data is explained by the first four PCA axes. Comparing the eigenvalues for the axes of the PCA with those expected under a broken stick model indicates that only the first principal component explains statistically significant amounts of the total variation. This axis explains 50% of the variation in the hydrochemical data. Table 5 shows the summary results for the analysis of the hydrochemical data.

Table 5: Summary statistics of the PCA of the hydrochemistry data.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.507	0.143	0.137	0.70	1.000
Cumulative % variance of species data	50.7	65.0	78.7	85.7	
Σ all unconstrained λ					1.000

Figure 14 shows the correlation biplot of the PCA ordination. In a correlation biplot, the plot is scaled so that the correlations between the explanatory variables, in this case the hydrochemical variables are emphasised. The explanatory variables are depicted as biplot arrows or vectors. In the correlation biplot, the plot is scaled in such a way that the angles

between vectors represent the correlations between explanatory variables. Small angles between vectors indicate close, positive correlations between explanatory variables, where as large (→ 180°) angles indicate strong, negative correlations between explanatory variables. Vectors that lie perpendicular to each other indicate that there is no correlation between those explanatory variables.

The most apparent aspect of this biplot (Figure 14) is that there are close correlations between groups of hydrochemical variables. For example, there are strong positive correlations between pH, alkalinity and calcium, all of which are closely correlated with the first axis if the PCA. Other hydrochemical variables are also closely correlated with this first axis and the acidity-related variables (e.g. conductivity, and major ions). The aluminium variables (Total, labile and monomeric) as one would expect are closely correlated with each other. Aluminium is correlated with both PCA axes 1 and 2, explaining variances in the samples along both the principal axes of variation in the ordination. Nitrate is also closely correlated with aluminium and negatively correlated with pH, which is to be expected as nitrate is a contributing factor to the acidification of surface waters. What is more surprising though, is that, across the 163 hydrochemical samples, there is little or no correlation between sulphate and nitrate, though both variables explain variances in samples on PCA axis 1 and 2. As both sulphate and nitrate are contributory factors to the acidity of aquatic environments, one would have expected that there be a degree of correlation between the two variables.

The PCA confirms the presence of two strong gradients in the hydrochemical data: an acidity-related gradient and an electrical conductivity gradient. The conductivity gradient is primarily the result of the presence of many lakes in the data set from North-western Scotland, where sea-salt events are known to influence the hydrochemistry of these lakes (Allott *et al.* 1995).

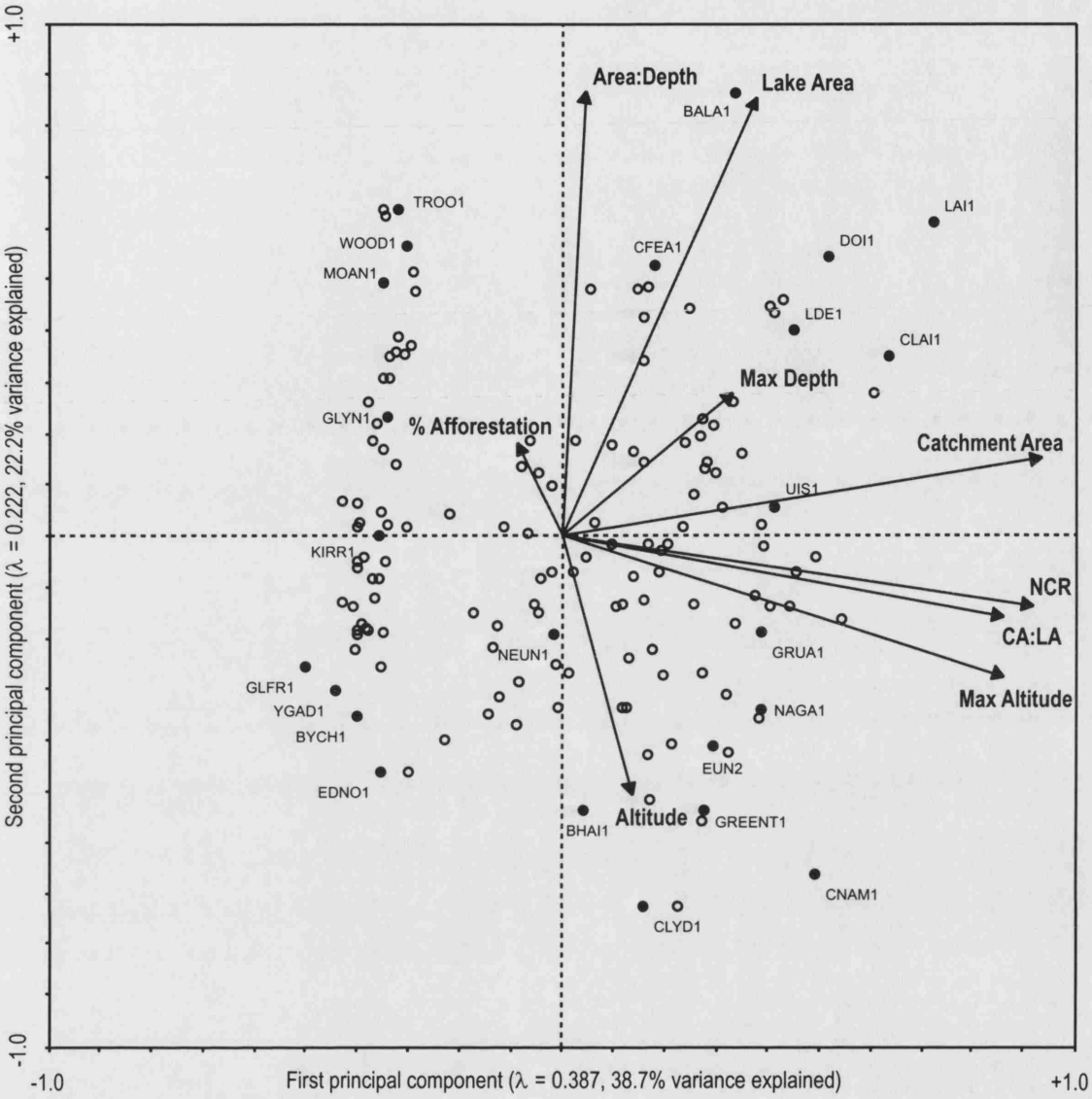


Figure 15: Correlation biplot showing the results of the principal components analysis of the 19-variable, 163-sample physical properties data set. The values in brackets are the eigenvalues for the respective principle components. To improve the clarity of the biplot only some of the samples are labelled (see appendix A for sample codes). Area:Depth = Lake area to depth ratio, NCR = net catchment relief, CA:LA = catchment area to lake area ratio. See text for interpretation

Table 6: Summary statistics of the PCA of the physical variables.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.387	0.222	0.131	0.108	1.000
Cumulative & variance of species data	38.7	60.9	73.9	84.7	
∑all unconstrained λ					1.000

The PCA of the physical variables (Figure 15, Table 6) also indicates the presence of two strong, independent gradients in the data. All variables in the physical data were $\log_{10}$ transformed prior to analysis, except maximum altitude (MaxAlt), and the replicate samples were given weight = 0.5 to reduce the effects of these replicates on the ordination. The summary results of the PCA are shown in Table 6. The first two axes of the PCA are significant under the broken stick model.

Figure 15 shows the correlation biplot of the ordination of the physical variable dataset. The first axis of the PCA is closely correlated with catchment area, net catchment relief, catchment area to lake area ratio and maximum altitude in the catchment. Lakes on the right of the plot are located in large catchments, which contain an above average amount of high ground. The second axis of the PCA is a contrast between lakes of large lakes located at lower altitudes found towards the top of the plot, and small, higher altitude lakes found towards the bottom. Although not plotted, the third axis of this PCA is strongly correlated with maximum lake depth and the ratio of the catchment area to the lake area.

The 24-variable, 163-sample physico-chemical data set was also analysed using PCA to investigate the relative explanatory power of the hydrochemical and the physical variables across the dataset (Figure 16). Again, the replicate samples were given weight = 0.5 in the analysis to reduce any unwanted influence on the ordination as a result of the presence of replicates in the data. Figure 16 shows that both the hydrochemical and the physical variables have similar levels of explanatory power in the ordination. The primary gradient (the first principle component) is an acidity and conductivity gradient. This indicates that the differences in the acidity and conductivity-related variables between samples across the dataset are the most important pattern in the lake data. Altitude is strongly negatively correlated with the first principal component, with the high altitude lakes being those that are of below average ionic strength and above average acidity. Figure 16 indicates that the main pattern in the lake physico-chemical data is a combination of both physical and hydrochemical variables, with both having similar levels of power in explaining the variation in the data.

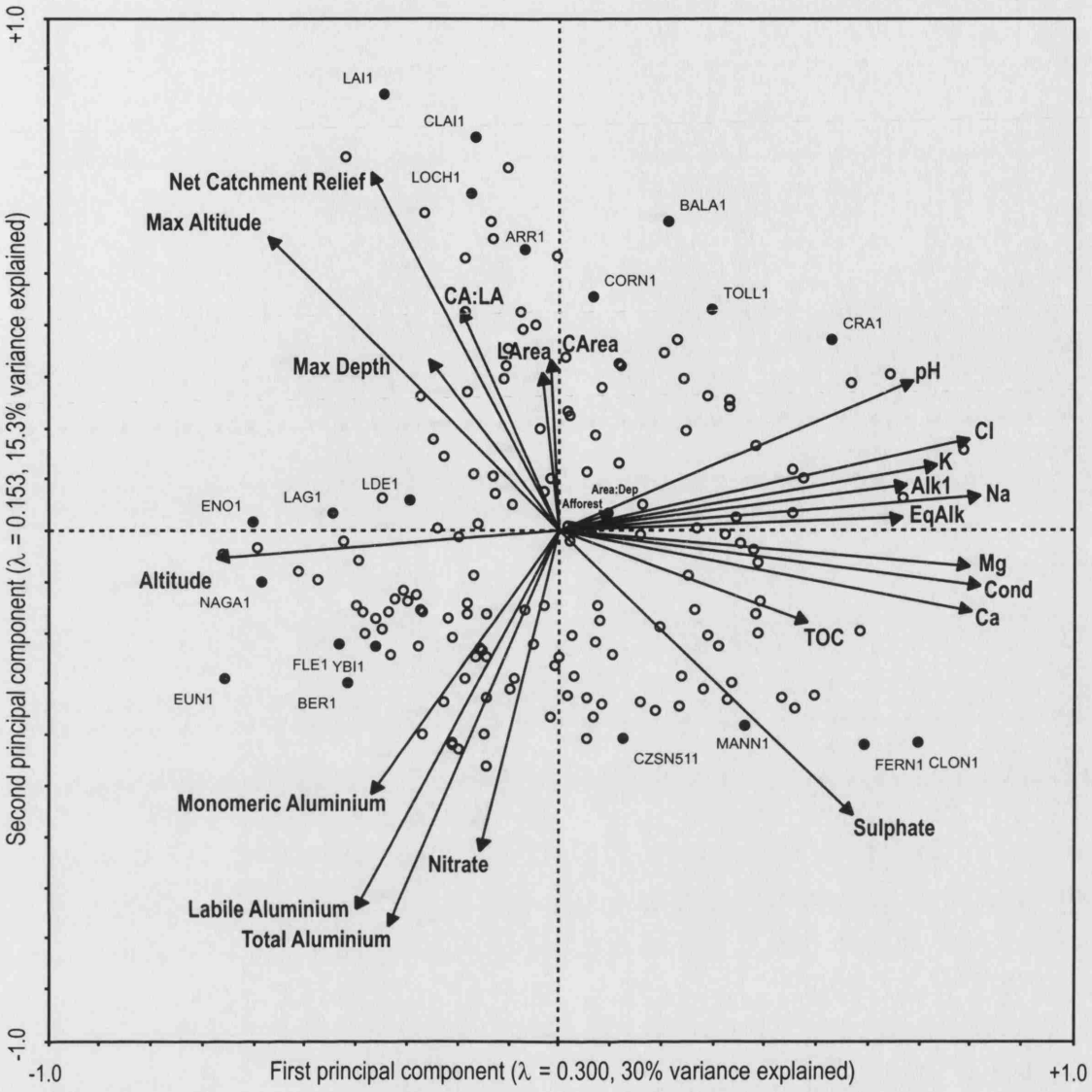


Figure 16: Correlation biplot showing the results of the principal components analysis of the 24-variable 163-sample physico-chemical dataset. The values in the brackets are the eigenvalues of the respective principal components. To improve the clarity of the biplot only some of the samples are labelled (see appendix A for sample codes). % afforestation (Afforest) and area to depth ratio (Area:Dep) have been plotted with smaller text to aid improve the clarity of the centre of the plot. CA:LA = catchment area to lake area ratio, LArea = lake area, CArea = catchment area, Cl = chloride, K = potassium, Alk1 = alkalinity, Na = Sodium, EqAlk = Equivalent alkalinity, Mg = magnesium, Cond = conductivity, Ca = calcium, TOC = total organic carbon.

Table 7: Summary statistics of the PCA of the physico-chemical data.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.304	0.161	0.177	0.083	1.000
Cumulative & variance of species data	30.4	46.5	58.2	66.6	
∑ all unconstrained λ					1.000

Axes three and four of this PCA are shown to be significant under the broken stick model, though they account for relatively small amounts of the total variance in the physico-chemical data, and little more variance than expected if the variance were randomly distributed amongst the principal components.

3.2.2 Surface sediment diatom data

The species data of the UKAWDDS are counts of diatoms, recovered from surface-sediment samples from each of the 151-lakes in the dataset. As already mentioned, 12 of the 151-lakes are represented by 2 separate samples taken in different years, bringing the total to 163-samples in the dataset. Each of the 163 diatom counts corresponds to a hydrochemistry sample in the physico-chemical dataset. The majority of the hydrochemical samples are averages of quarterly (or more) water samples, and therefore correspond to the mean hydrochemistry for the period that the diatoms are meant to represent.

The species have been harmonised to conform to the taxonomy used in the SWAP training set (Stevenson *et al.* 1991). This harmonisation was required because those samples counted since the SWAP training set was published have benefited from changes in taxonomy and changes in nomenclature. Furthermore, decisions were made during the SWAP palaeoecology programme to combine taxa due to problems differentiating between varieties or morphotypes of species.

The diatom data contain 394 individual species, though many are rare taxa occurring either in very low abundances or only in a few samples. Unless otherwise stated in the text, the diatom data analysed in this, and subsequent chapters, contain only those taxa that occur with minimum percentage abundance $\geq 2\%$ in any one sample. After removing those taxa that do not conform to the minimum abundance rule, the diatom data contained 189 taxa. There are a number of reasons for this deletion. Firstly, many of the methods employed do not perform optimally where the data contain many rare taxa. Secondly, one could argue that these rare taxa represent natural noise in the data and can safely be ignored. On the other hand they are important biological indicators and should be considered, therefore, in any analysis. However, many of the methods considered in this and subsequent chapters are designed to extract the main patterns of variation in the

multivariate data. In this case, the removal of the rare taxa should not influence the interpretation of the main patterns in the data.

To determine the gradient lengths in the diatom data, and thus choose an appropriate ordination technique with which to analyse it, a detrended correspondence analysis (DCA) was performed with detrending by segments, non-linear rescaling, and Hill's biplot scaling of the species and sample scores. The resulting ordination has the useful property of expressing the units of the axes of the DCA in standard deviation (s.d.) units and is, therefore, an expression of species turnover along the axes of the ordination. Species abundance rises to a maximum and then declines again in approximately 4 s.d, so that samples that lie 4 s.d. or more apart have no species in common. Gradient lengths longer than *ca.* 2 s.d. suggest a unimodal response of the species to the underlying environmental gradients in the data. In such cases unimodal-based analysis techniques would be appropriate.

The DCA of the diatom data returned a gradient length of 3.446 s.d. units for the first axis and 2.192 for the second. The third and fourth axes of the ordination showed gradient lengths of 1.9 and 1.8 s.d units respectively. The variance explained by each of the four axes of the DCA is significant when compared to the variance expected under the broken stick distribution. These results indicate that the response of the diatoms to the underlying environmental gradients in the dataset is unimodal. Therefore, CA was used to extract the main patterns of variance in the diatom data (Table 8).

The CA of the diatom data showed a prominent arch in the pattern of both the species and the site scatterplots. Furthermore, the eigenvalue of the second axis of the CA is less than half the eigenvalue of the first axis. In circumstances where this is the case, the arch effect is likely to be a mathematical artefact of the method, and not a property of, or underlying structure in, the diatom data (Hill and Gauch, Jr. 1980). The arch effect is a known problem of CA and has been discussed above (see section 2.4.1.2 above). Methods for detrending the arch effect are available, including detrending by segments (with non-linear rescaling) or by polynomials of second, third, or fourth order. DCA is the ordination technique that implements a detrending of the arch.

Table 8: Summary statistics of the CA of the diatom data.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.364	0.155	0.133	0.124	3.147
Cumulative & variance of species data	11.6	16.5	20.7	24.7	
∑ all unconstrained λ					3.147

DCA was used, therefore, to extract the major patterns in the diatom dataset. Detrending by segments with non-linear rescaling was used because detrending by polynomials, whilst being a less severe method of detrending, provided little improvement over detrending by segments. In addition, non-linear rescaling is not an option when detrending by polynomials, so detrending by segments allows the structure at the ends of the first axis of the ordination to be extracted more usefully.

The results of the Dca of the diatom data are shown in Table 9, Figure 17 and Figure 18. The first axis of the DCA explains 12.2% of the variance in the diatom data (λ_1 = 0.365, total inertia = 2.993). A further 11.4% of the variance in the diatom data is explained by axes two to four. All four of the DCA axes are significant when their eigenvalues are compared to the broken stick distribution, and explain 23.6% of the variance in the data. The results of the DCA indicate that there is a single major axis of variation in the diatom data. This axis is strongly correlated (Pearson product moment correlation coefficient) with pH (r = 0.824), calcium (r = 0.762), equivalent alkalinity (r = 0.695), and total and monomeric aluminium (r = -0.683 and r= -0.684 respectively).

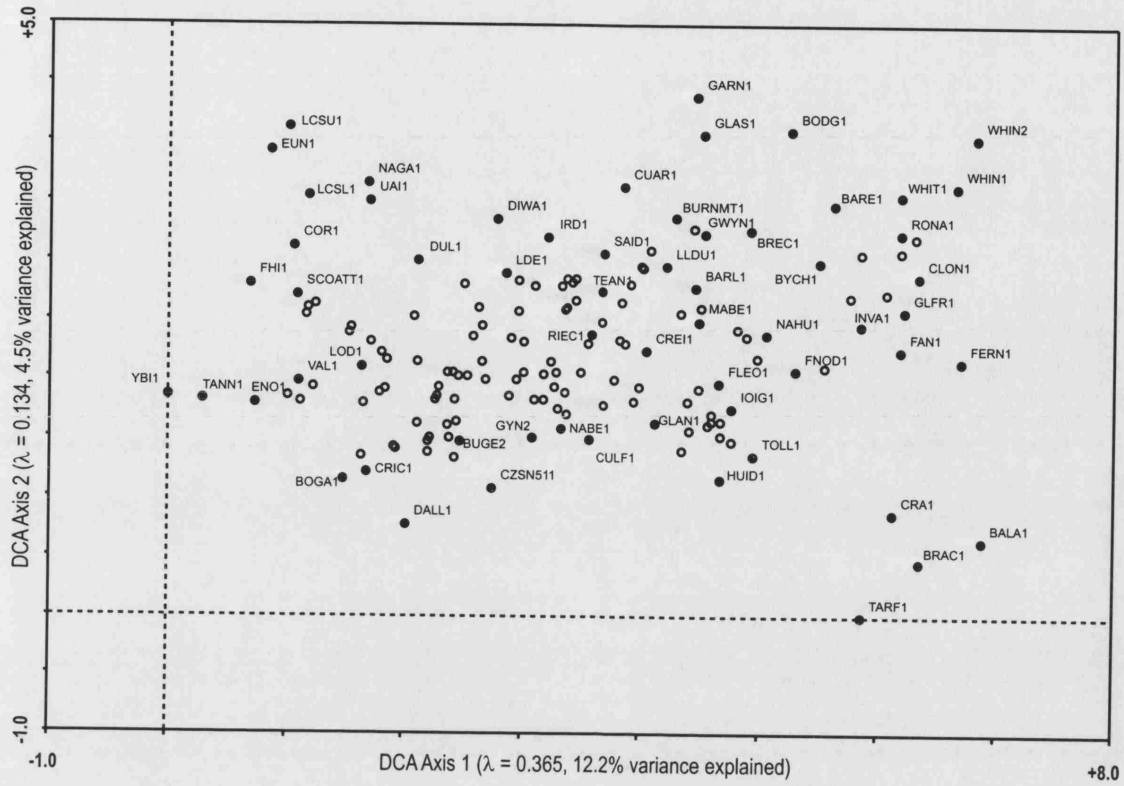


Figure 17: Plot of the site scores on the first and second axes of the DCA of the diatom species data. The numbers in brackets are the eigenvalues of the analysis, a reflection of the amount of the total variance explained by the axis. Total variance in the diatom data (inertia) = 2.993. To improve the clarity of the plot, not all points are labelled. Labels are the sample codes, see appendix B for explanation of these codes.

The biplot of the site scores (Figure 17) shows that the most acid lakes are located towards the left of the plot (e.g. Llyn y Bi [YBI1, pH = 4.92], Loch Enoch [ENO1, pH = 4.54]). These lakes are associated with low pH and acid buffering capacity. Those lakes with a higher pH (e.g. Llyn Tegid [BALA1, pH = 6.36], Loch Fern [FERN1, pH = 6.84] and Loch Whinyeon [WHIN1, pH = 0.692]) are located on the left of the plot. These lakes are associated with circumneutral pH and higher acid neutralising capacities.

Table 9: Summary statistics of the DCA of the diatom data.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.364	0.128	0.112	0.080	3.147
Gradient length (s.d.)	3.468	1.886	1.837	2.199	
Cumulative & variance of species data	11.6	15.6	19.2	21.7	
∑ all unconstrained λ					3.147

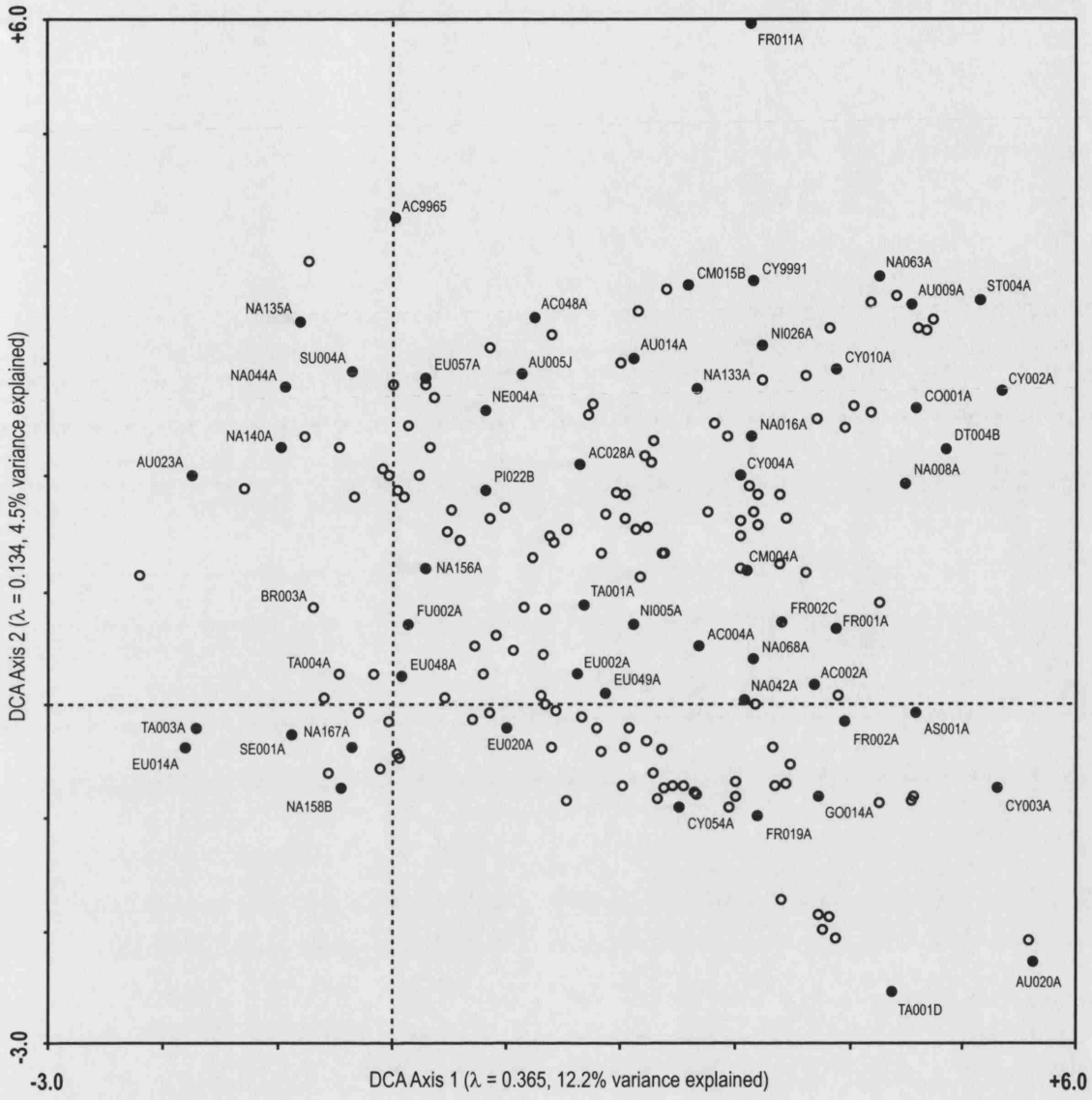


Figure 18: Plot of the species scores on the first and second axes of the DCA of the diatom species data. The numbers in brackets are the eigenvalues of the analysis, a reflection of the amount of the total variance explained by the axis. Total variance in the diatom data (inertia) = 2.993. To improve the clarity of the plot, not all points are labelled. Labels are the species codes, see appendix B for explanation of these codes.

The other major pattern in the hydrochemistry data, the conductivity gradient, is not reflected to such a degree in the diatom data as the acidity gradient, which plays a more dominant role in determining species composition across the 163 samples.

The ordering of the sites on the left of the DCA site plot does not reflect the ordering of the sites if we were to consider pH alone. pH perhaps acts in a direction that is at a slight angle to the first DCA axis, and this is reflected in the Pearson product moment correlation coefficient between pH and the DCA axis 1 site scores (r = 0.824). The other acidity-related variables (e.g. aluminium, calcium and alkalinity) may also account for variance along the first axis of the DCA. This may explain the ordering of the lakes along this axis.

The second axis of the DCA is difficult to interpret, not only because of the detrending process, but also because of the complex gradients in the physico-chemical data. Pearson product moment correlations between the DCA axis 2 site scores and the chemistry data indicates that there are linear correlations between axis 2 scores and total organic carbon (r = -0.585), calcium (r = -0.515), conductivity (r = -0.484) and monomeric aluminium (r = 0.463).

The acidity-related gradient of DCA axis 1 is further reflected in the positions of the species in the DCA species plot (Figure 18). Acid tolerant taxa (e.g. *Eunotia bactriana* [EU014A, SWAP pH optimum = 4.7], *Tabellaria binalis* [TA003A, SWAP pH optimum = 4.7], and *Tabellaria quadriseptata* [TA004A, SWAP pH optimum = 4.9]) are located to the left of the species plot. Acid-sensitive diatoms (e.g. *Aulacoseira islandica* [AU009A, SWAP pH optimum = 6.5], *Cyclotella pseudostelligera* [CY002A, SWAP pH optimum = 6.9], and *Cyclotella meneghiniana* [CY003A, SWAP pH optimum = 6.9]) are located on the right of the plot. Again, the second axis of the DCA is difficult to interpret. The patterns in the diatom data will be discussed further below.

3.2.3 Diatom responses to the physico-chemical variables

Diatom responses to the physico-chemical data were analysed using canonical correspondence analysis (CCA). CCA was chosen to analyses the data after a DCCA showed that the diatom data exhibit unimodal responses to the extracted environmental axes (Axis 1 gradient length = 2.810 s.d.).

The biplot of the species-environment relationships when all 24 physico-chemical variables were included in the ordination showed a strong arch in the plotted site and species points. CCA is known to suffer from the arch effect in the same manner as CA because CCA is just a constrained form of correspondence analysis. Arch effects tend to occur in CCA where superfluous environmental variables are included in the ordination. It is desirable to remove these superfluous environmental variables before using DCCA to extract species-environment relationships. To this end, a forward selection procedure was performed using the CANOCO software in order to reduce the number of environmental variables used in the analysis.

Forward selection was used to select the minimum set of environmental variables that explained statistically significant amounts of the variance in the species data. A Bonferroni correction to the p-value, used to determine the significance of the environmental variables, was made. This correction is needed because multiple tests are being performed simultaneously. If the correction was not applied then the chances of accepting a result as being significant when this was not the case increases as more simultaneous tests are performed. The p-value is adjusted so that the required p-value (p') ($\alpha = 0.05$) is equal to the original p-value (p) divided by the number of tests being run. The results of the forward selection are shown in Table 10.

Table 10: Results of the forward selection of the 24 physico-chemical variables. Only those variables that explain statistically significant additional amounts of variance in the diatom data set are included. Statistical significance assessed at α=0.05 using 999 Monte Carlo permutations tests. P-value is the significance of the F-value assessed by permutation tests. P-required is a Bonferroni corrected P-value. λ = eigenvalue for the environmental variable and is a measure of the (extra) variance explained. % is the extra variance explained expressed as a percentage.

Variable	F	P-value	P-required	λ	%
pH	16.108	0.001	0.05	0.272	9.1
Total organic carbon (TOC)	5.072	0.001	0.025	0.084	2.8
Labile Aluminium (Allab)	5.036	0.001	0.017	0.081	2.7
Max lake depth (MLDepth)	3.251	0.001	0.0125	0.052	1.7
Calcium (Ca)	2.773	0.001	0.01	0.043	1.4
Lake Altitude (LAlt)	2.569	0.001	0.0083	0.010	1.3
Potassium (K)	2.305	0.001	0.007	0.035	1.2

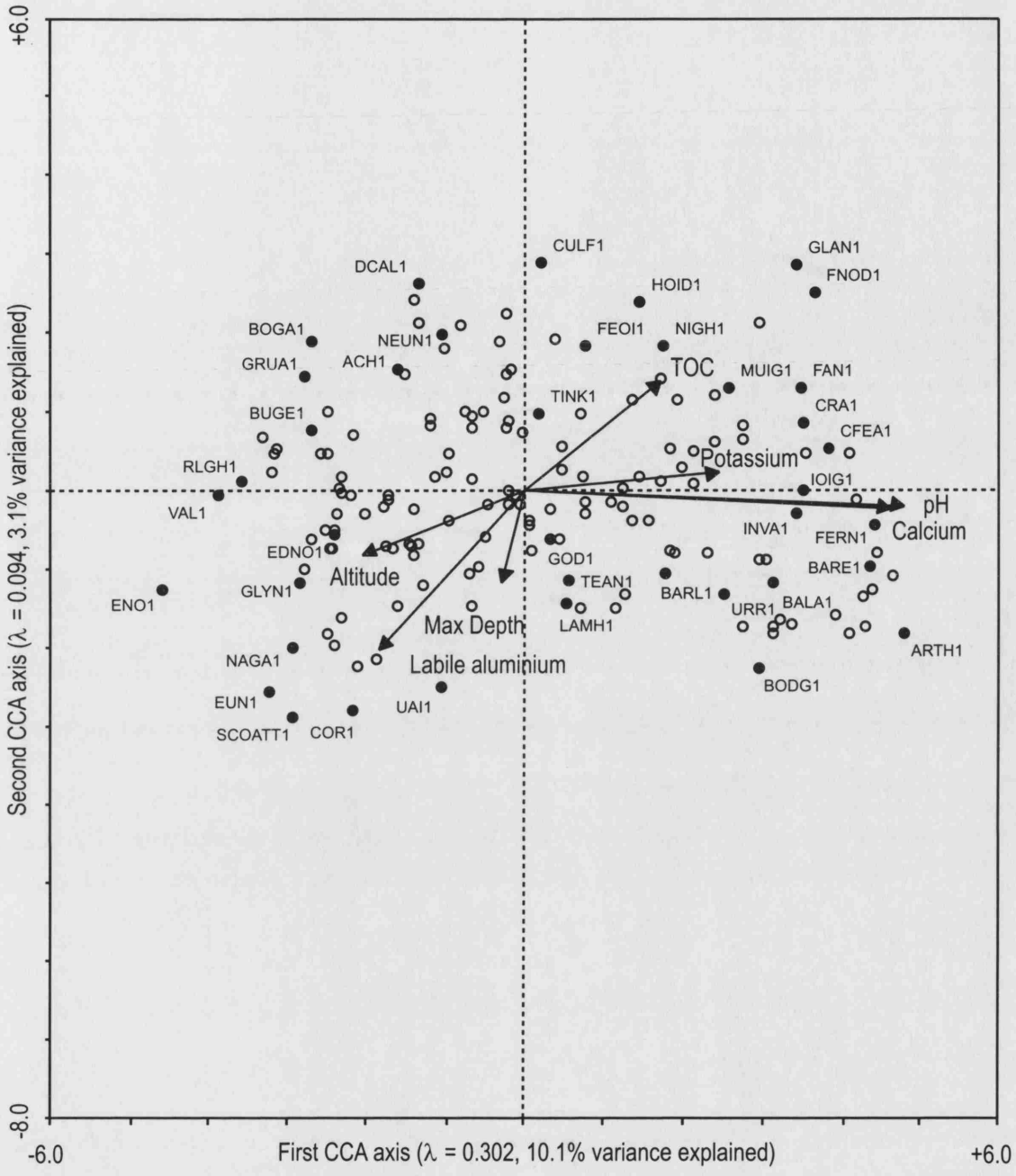


Figure 19: CCA biplot showing a plot of the site scores from the analysis and the seven significant environmental variables retained for analysis. The figures in brackets are the eigenvalues for the axes plotted. Total inertia = 2.993.

The seven variables chosen in the forward selection procedure were then used as environmental variables in a CCA of the diatom species data. The removal of the extra environmental variables resulted in the removal of the arch in the biplot of the site scores from the CCA. The biplot of the CCA species scores on CCA axes one and two, however, still showed the presence of a slight arch in the species environment biplot. This arch,

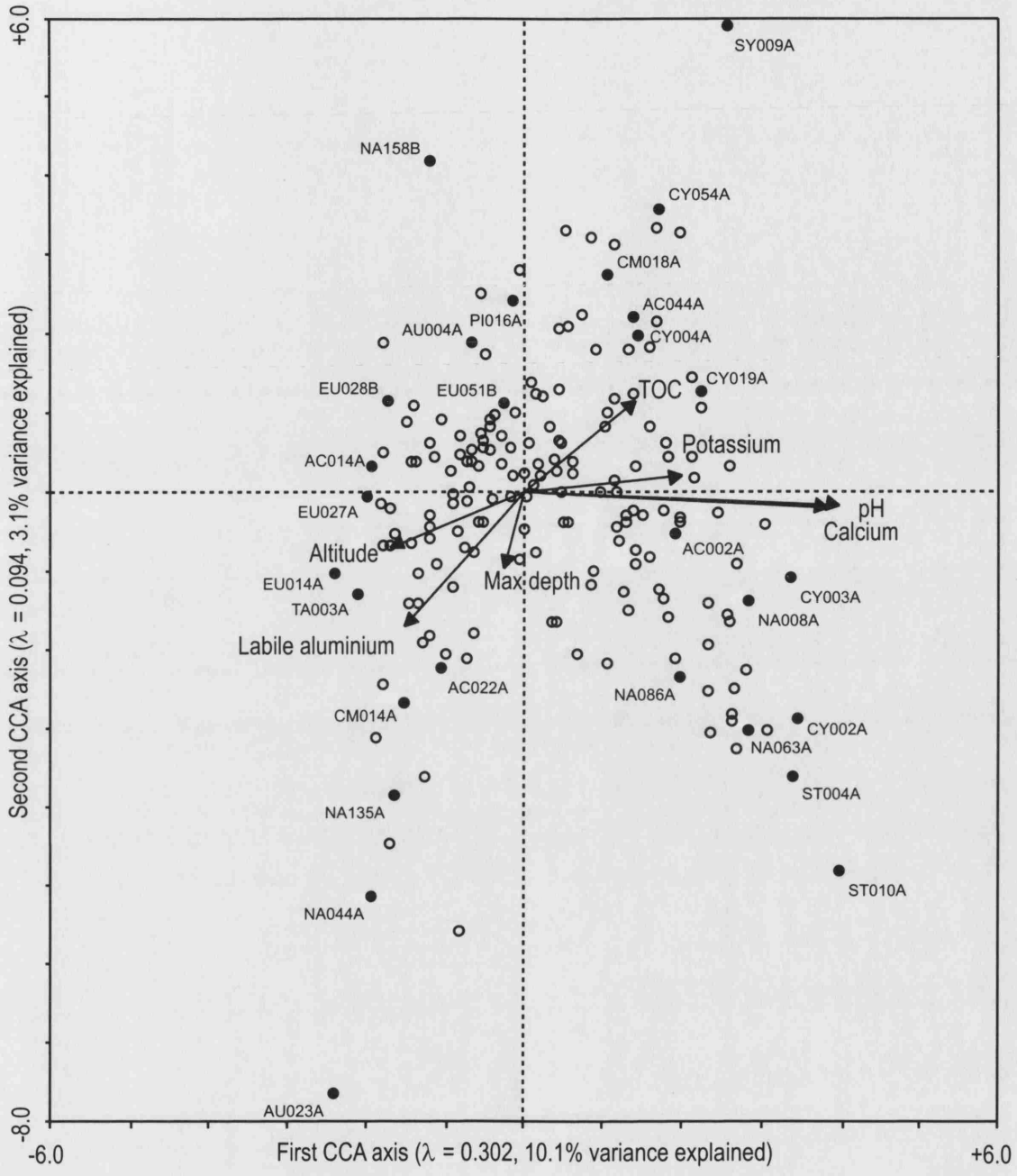


Figure 20: CCA biplot showing a plot of the species scores from the analysis and the seven significant environmental variables retained for analysis. The figures in brackets are the eigenvalues for the axes plotted. Total inertia = 2.993.

however, is more likely to be the result of the interrelationship between pH and aluminium and the way in which these variables are portrayed in the 2-dimensional space described by the biplot.

The arch arises when the algorithm cannot find a good second axis that is uncorrelated with the first, and so it folds the first axis and takes this as the second axis. This folded axis

has the property of being uncorrelated with the first. If this were the case in this CCA, pH should also be correlated with the second axis as well as the first. Furthermore, the variance inflation factors for each of the nine physico-chemical variables indicate that none of the variables is largely explained by other variables in the ordination. The arch is likely to be the result of the correlation between pH and aluminium, and because aluminium explains a unique amount of variance in the diatom data, not the result of an artefact of the method.

A DCCA of the diatom and physico-chemical data was analysed using second order polynomials to detrend the ordination. When this ordination is compared to the CCA (seven variables), the CCA has greater explanatory power than the DCCA. Furthermore, DCCA is a complex solution to the problem of ordination. It makes many assumptions about the data and the mathematical theory behind the method has yet to be robustly tested. In the interests of parsimony, one should try to use CCA before DCCA wherever possible. The use of CCA is justified, therefore, in this case.

The results of the CCA (seven physico-chemical variables) are shown in Table 11. The first CCA axis explains 9.8% of the species data, with the second and subsequent axes explaining 3.3%, 2.3% and 1.8% respectively. The seven statistically significant physico-chemical variables explain 21.96% of the variance in the diatom data. The first axis is the dominant axis of variation, explaining three times the amount of variance explained by the second or subsequent axes.

Table 11: Summary results of the CCA of the diatom data and the seven statistically significant variables chosen using forward selection.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.302	0.094	0.065	0.047	2.993
Species environment correlation	0.919	0.808	0.763	0.726	
Cumulative % variance of species data	10.1	13.2	15.4	17.0	
Cumulative % variance of species environment correlation	49.8	65.3	75.9	83.7	
∑ all unconstrained λ					2.993
∑ all canonical λ					0.607

The site and physico-chemical variable biplot is shown in Figure 19. The ordination shows the clear dominance of the acidity gradient in the data set. The first axis is a contrast between those lakes with high pH and calcium, found to the right of the plot, and those lakes with low pH and calcium on the left. pH and calcium are strongly correlated, indicated by the small angles between the respective biplot arrows. Calcium is usually found to be covariant with pH and independently explains little of the total variance in the diatom data (1.3%). Aluminium is, as expected, also correlated with the first axis of the CCA. Aluminium is also correlated with CCA axis 2, and independently explains and extra 2.8% of the variance in the diatom data.

The most acid lakes (e.g. Loch Enoch [ENO1, pH = 4.54], Round Loch of Glenhead [RLGH1, pH = 4.73] and Lochnagar [NAGA1, pH = 5.05]) are located on the left of the biplot because of the dominance of acid-tolerant taxa in these lakes (shown in Figure 19). Lakes that have more circumneutral waters (e.g. Loch Arthur [ARTH1, pH = 7.17], Loch nam Brac [BRAC1, pH = 6.99], Loch na Claise Ferna[CFEA1, pH = 6.99]) are found to the right of the plot. The diatom assemblages from the acid lakes are dominated by acid tolerant taxa (e.g. *T. binalis* [TA003A, SWAP pH optima = 4.7] and *E. bactriana* [EU014A, SWAP pH optima = 4.7]) as shown in the species-environment joint plot (Figure 20). These lakes also have few acid sensitive taxa, shown by the large distances between the position of the acid lakes and the positions of the acid sensitive diatoms. In the biplot, the species are fitted in such a way the abundance of taxon *i* declines in all directions away from the point for taxon *i* on the biplot.

The second axis of the CCA is a complex environmental gradient. As well as labile aluminium (Allab) being correlated with axis 2, total organic carbon (TOC), altitude (MaxAlt) and net catchment relief (NCR), and max lake depth (MLDepth) are also closely correlated with this axis. TOC is negatively correlated with the other variables. NCR and MLDepth are closely correlated with CCA axis 2 and explain little of the variance along axis 1. Allab, TOC and MaxAlt are correlated with both CCA axes 1 and 2. The second axis is probably a contrast between higher altitude lakes where catchment soils are perhaps less well developed and contribute little allochthonous organic matter to the lakes, and lower altitude lakes, draining well developed peaty soils leading to high levels of TOC and humic acids in the water column. The negative correlation between labile Aluminium and TOC is

a reflection of the decreased mobility of aluminium as it binds with TOC in the water column.

3.3 Comparison of the UKAWDDS and the 83-lake diatom training set

Of the 163 samples in the UKAWDDS only 83 of these had sufficient sediment remaining to be used in a cladoceran preparation. As such it is important to take this subset of the main UKAWDDS and check to see whether similar patterns in the diatom-physico-chemistry responses identified in the UKAWDDS also exist in this subset.

This section of the thesis will compare the 83 lake subset with the UKAWDDS using ordination techniques. The analysis of the physico-chemical data for the 83-lake subset is presented in chapter 4.

Long gradients in the diatom data for the 83-lake data set, as indicated by detrended correspondence analysis of the data, indicate that unimodal ordination methods would be suitable to determine the latent structure in the species information. Correspondence analysis, with square root transformation of the species data and down weighting of the rare taxa, was performed on the diatom data for the 83-lake training set. The results of this CA indicate the presence of a single, primary gradient in the underlying structure of the diatom species information. The three subsequently extracted axes of the CA are also shown significant when the variance they explain is compared to the variance expected under the broken stick distribution.

The first axis of the CA explains 12.6% of the variance in the diatom data ($\lambda_1 = 0.338$, total inertia = 2.687), with the three subsequent axes explaining 7.4%, 6.2% and 5.2% respectively ($\lambda_2 = 0.174$, $\lambda_3 = 0.167$, and $\lambda_4 = 0.137$), for a total of 30.4% of the total variance in the species data explained.

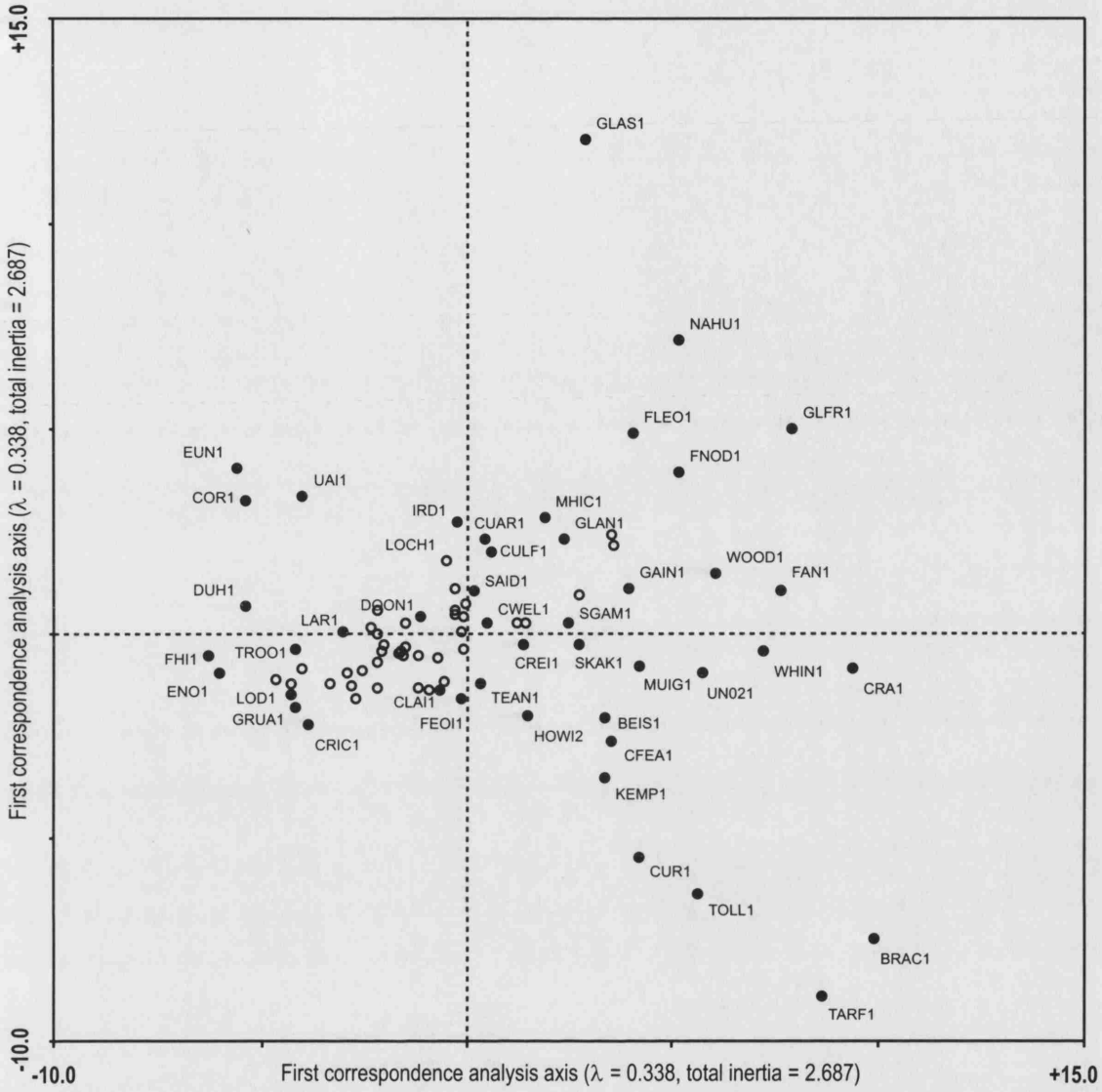


Figure 21: Correspondence analysis biplot of the site scores for the 83-lake training set, based on the diatom surface sediment data.

Correlations between the physico-chemical data and the sample scores of the CA of the diatom data indicate that the first axis of the correspondence analysis is strongly associated with a gradient of acidity. pH ($r = 0.828$), calcium, ($r = 0.733$), equivalent alkalinity ($r = 0.685$) and aluminium (total, $r = -0.627$; labile, $r = -0.626$; and monomeric, $r = -0.719$) are all strongly correlated with the first axis of the CA.

The second axis of the correspondence analysis is not particularly correlated with any of the 24 physico-chemical variables measured across the 83-lake training set. The closest correlations between the measured physico-chemical variables and the samples scores on

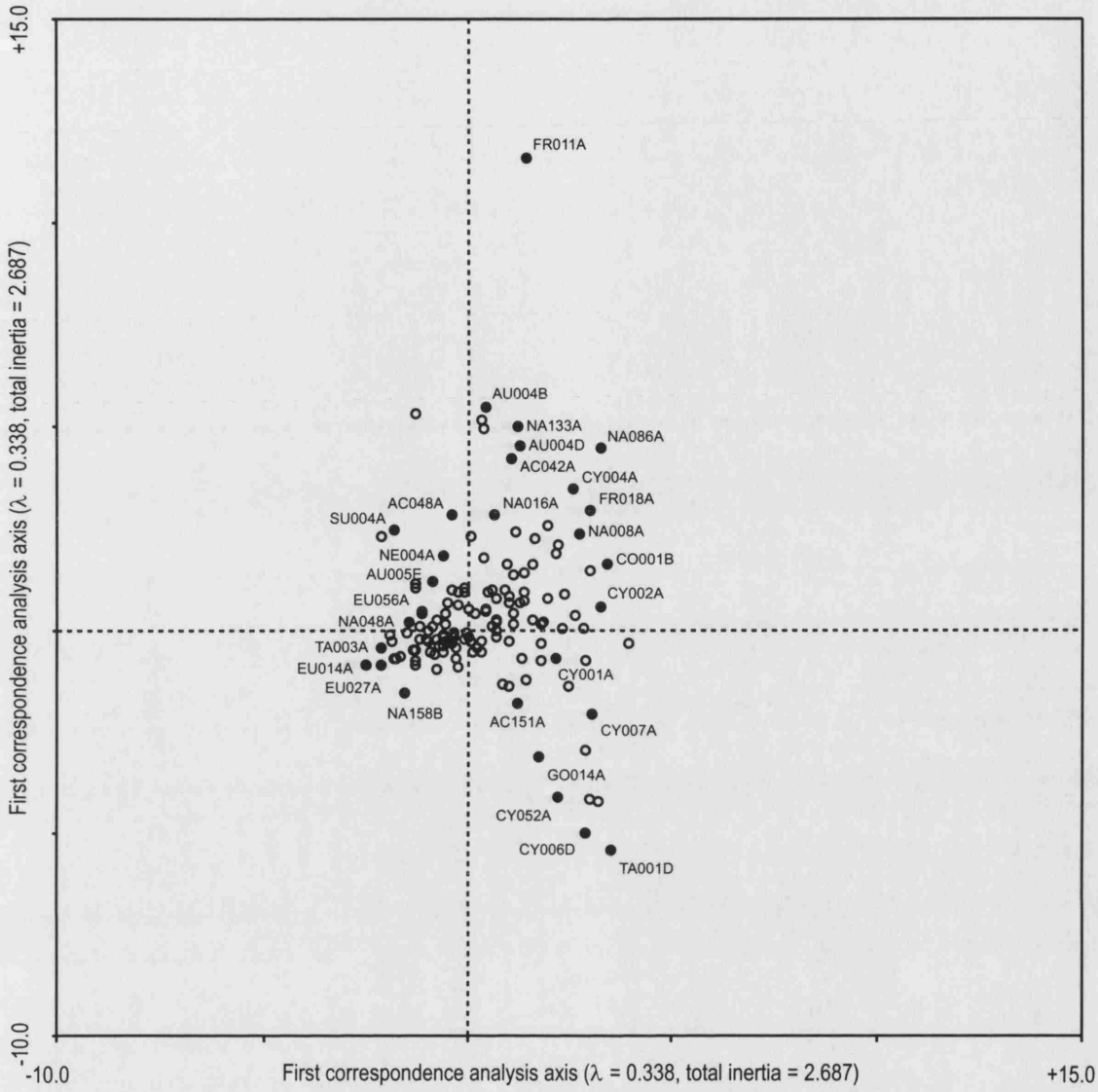


Figure 22: Correspondence analysis biplot of the species scores for the 83-lake training set, based on the diatom surface sediment data.

axis two of the CA are for maximum altitude in the catchment (r = 0.383), and nitrate (r = 0.333).

Axis three of the CA is associated with the aluminium concentrations in the lakes (labile, r = 0.644; monomeric, r = 0.517; and total, r = 0.554), the amount of organic matter or organic acidity (total organic carbon, r = −0.707), and maximum depth (r= 0.441). The fourth axis of the correspondence analysis, like the second, is uncorrelated with the 24 measured physico-chemical variables.

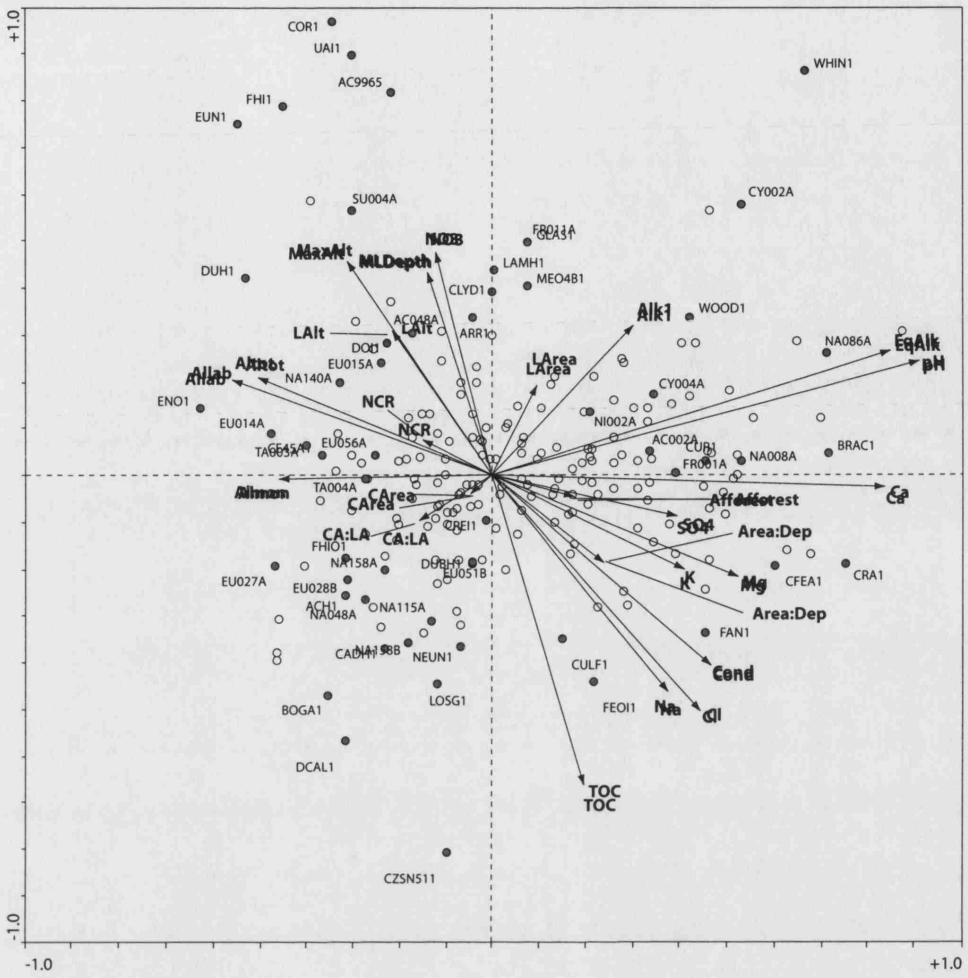


Figure 23: CCA biplot showing the relationship between the species scores and the physicochemical variables in the analysis of the 83-lake diatom training set.

Constrained ordination of the 83-lake diatom data set was performed using CCA after DCCA showed that the gradient lengths in the diatom data with respect to the physico-chemical parameters were quite long (axis 1 = 2.955 s.d.). The results of the CCA of the 83-lake diatom training set are shown in Figure 24, Figure 23, Table 12 and Table 13

The main axis of variation in the 83-lake diatom data is an acidity gradient as indicated by the long biplot arrows and inter set correlations for pH (inter set correlation with CCA axis 1 = 0.8397), equivalent alkalinity (0.7803) and calcium (0.7714).

It is not appropriate in this case to use the regression coefficients and their associated t-values to assess which variables are significant in explaining variance on the CCA axes, because the 24 physico-chemical variables show a large of degree of multicolinearity. In

such instances the regression coefficients become unstable and the inter set correlations should be used instead (ter Braak and Smilauer 2002).

Table 12: Summary results for the CCA of the 83-lake diatom training set.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.293	0.122	0.104	0.089	2.687
Species environment correlation	0.940	0.882	0.828	0.866	
Cumulative % variance of species data	10.9	15.4	19.3	22.6	
Cumulative % variance of species environment correlation	25.3	35.9	44.9	52.6	
∑ all unconstrained λ					2.687
∑ all canonical λ					1.155

Table 13: Results of the forward selection of the 24 physico-chemical variables. Only those variables that explain statistically significant additional amounts of variance in the83-lake diatom data set. Statistical significance assessed at a=0.05 using 999 Monte Carlo permutations tests. P-value is the significance of the F-value assessed by permutation tests. P-required is a Bonferroni corrected P-value. λ = eigenvalue for the environmental variable and is a measure of the (extra) variance explained. % is the extra variance explained expressed as a percentage.

Variable	F	P-value	P-required	λ	%
pH	8.184	0.001	0.05	0.247	9.2
Maximum lake depth (MLDepth)	2.977	0.001	0.025	0.088	3.3
Maximum altitude in catchment (MaxAlt)	2.776	0.001	0.017	0.080	3.0
Net catchment relief (NCR)	2.582	0.001	0.0125	0.073	2.7
Total organic carbon (TOC)	1.873	0.002	0.01	0.052	1.9
Monomeric aluminium (Almon)	1.789	0.002	0.0083	0.049	1.8
Sulphate (SO4)	1.683	0.004	0.007	0.046	1.7
Totals				0.635	23.6

The biplots also indicate that aluminium is negatively correlated with this first CCA axis. The influence of aluminium on the diatom distributions in the 83-lake training set is much more related to the correlation between aluminium and pH than in the UKAWDDS. Aluminium was correlated closely with the second CCA axis in the UKAWDDS analysis but in the CCA of the 83-lake diatom data there correlation is much less strong, indicated by

the large angles between the three aluminium biplot arrows and the second CCA axis (see Figure 24 and Figure 23). This is also illustrated in the results of the forward selection procedure (Table 13), which shows a much reduced effect of aluminium on the diatom distributions independent of the other variables of 1.8% compared to 2.7% in the UKAWDDS. Whilst these figures seem low, all 24 physico-chemical variables can explain just 43% of the species data in the reduced training set because random noise in the training set is difficult to explain completely.

Whilst the overall configuration of biplot arrows and sites is similar in both the CCA results for the UKAWDDS and the CCA results for the reduced training set, forward selection of the physico-chemical variables does suggest some differences in the importance of certain variables in explaining the distributions of the diatom taxa in the two training sets (c.f. Table 10 and Table 13).

pH is the variable that explains the greatest amount of variance in both of the training sets (*ca.* 9%) and indicates that both training sets contain a similar pattern in the diatom relationship to pH. Altitude variables (MaxAlt and NCR in the 83-lake data set and LAlt in the UKAWDDS) explain more of the variance in the 83-lake training set compared to the UKAWDDS, whilst TOC and aluminium are more important in the UKAWDDS. The independent effect of potassium on the distributions of the diatom taxa in the UKAWDDS is no longer significant in the 83-lake training set, where sulphate concentrations are now more important.

It is difficult to explain to such small independent effects of variables in terms of ecological terms and it is perhaps unwise to draw too great a significance from these subtle differences in the two training sets. The biplots of the CCA of both data sets show similar configurations in the biplot arrows and the primary axis of variance in both is an acidity gradient. Some of the subtle variation in the important variables such as aluminium and calcium at specific values of pH seems to have been lost from the smaller data set as the result of reducing the number of sites. This is reflected in the results of the forward selection and the relative angles between biplot arrows in the analyses for the two training sets, with the independent effects of aluminium, TOC and calcium being reduced.

3.4 A comparison of the UKAWDDS with the SWAP calibration set

The Palaeolimnology Programme of the Surface Waters Acidification Project (SWAP) generated a surface-sediment diatom training set for 138 acid sensitive surface waters from England, Scotland, Wales, Sweden and Norway. Numerical analysis of this training set illustrated the strong relationship between lake water acidity and diatom species composition and subsequently the diatom-pH relationship derived from this training set was used to reconstruct the past pH lakes from their fossil diatom communities.

The SWAP calibration set, as the training set is known, was created by the amalgamation of a series of local training sets selected so that the SWAP calibration set represented as wide a range of water chemistry, catchment sensitivity and pollution loadings. The SWAP calibration set was designed to capture an acidity gradient whilst trying to minimise the importance of other environmental gradients. As such the calibration set covers a wide range of lake water acidity but is biased towards sites at the acid end of the range. The SWAP calibration set has a modal pH of 5.0 (mean pH = 5.59, median pH = 5.40) and 25% of the lakes in the training set have a pH of less than 5.

The SWAP calibration set was particularly good at capturing the diatom-pH relationship; however, the training set might under-estimate the importance of other environmental gradients in determining diatom species composition because of the sampling protocol. The UKAWDDS has been collated from a wider range of UK lakes, and unlike the SWAP calibration set no attempt has been made to restrict the environmental gradients covered by the UKAWDDS training set.

In previous sections the UKAWDDS has been analysed numerically to identify the main patterns in the data set. It is important to compare the SWAP calibration set and the UKAWDDS in order to evaluate the consistency of results from these training sets. This section of the thesis will compare the results of the SWAP calibration with the results presented previously for the UKAWDDS. The two data sets contain information of a common core set of ten determinands. As well as comparing the published results of the SWAP calibration set with those of the UKAWDDS, a new analysis of the two training sets based on this core of common determinands will be presented.

3.4.1 Water Chemistry

Principal components analysis of the SWAP calibration set showed that the main pattern in the chemistry data was the strong gradient from high to low pH, alkalinity, conductivity and calcium. The second axis of this PCA was a complex gradient, which correlated well with aluminium concentrations, as well as those variables strongly correlated with the first axis. Lake altitude and maximum depth were also strongly correlated with both the first and second axes of the PCA.

The overall pattern shown in the PCA of the SWAP calibration set is reflected in the analysis of the UKAWDDS. This is not surprising as the two data sets are restricted to acid-sensitive and oligotrophic upland surface waters. There are a number of subtle differences between the two data sets however.

The first axis of the PCA of the UKAWDDS is a strong gradient from high to low pH, alkalinity, conductivity and calcium, similar to that shown in the SWAP calibration set. The UKAWDDS differs from the SWAP calibration set in that these variables are very strongly correlated with the first PCA axis and are almost uncorrelated with the second PCA axis. The UKAWDDS is clearly dominated by a strong acidity and conductivity gradient in much the same way as the SWAP calibration set; however the other main underlying gradient in the UKAWDDS is uncorrelated with these acidity and conductivity variables.

Altitude is also strongly correlated with the first PCA axis of the UKAWDDS, but unlike the SWAP calibration set altitude is largely uncorrelated with second PCA axis in the UKAWDDS.

The second axis of the PCA in both training sets is correlated with aluminium concentrations, maximum depth and lake altitude. The secondary gradient in the UKAWDDS however, is much more independent of the primary gradient than that in the SWAP calibration set. The secondary gradient in the UKAWDDS is much more closely associated with aluminium, maximum depth and altitude alone and is not correlated with the acidity and conductivity gradients that dominate the data set. The secondary gradient in the SWAP calibration set is a complex of mix of acidity, aluminium, conductivity altitude and depth.

A further difference between the two data sets is that in the SWAP calibration set altitude and maximum depth are negatively correlated with aluminium concentrations. In the UKAWDDS however, aluminium correlations are uncorrelated with altitude and maximum depth.

3.4.2 Water chemistry and surface-sediment diatom relationships

CCA was used to analyse the diatom-environment relationships in both the SWAP and the UKAWDDS training sets. The analysis of the SWAP calibration set used sixteen environmental variables, whilst the analysis of the UKAWDDS included 24 physico-chemical variables. A greater number of hydrochemical determinands were present for the samples in the UKAWDDS, but those determinands present in the SWAP calibration set were also present in the UKAWDDS. A number of the physical variables included in the SWAP analysis were associated with the amount of forestry in the catchment and the proportions of deciduous and coniferous forest also recorded. In the UKAWDDS only the amount of forest in the catchment was consistently available across all 163 samples. The UKAWDDS contains further information on lake area and catchment relief, variables not included for analysis in the original SWAP programme.

Despite these differences in the number and type of physico-chemical variables included in the analysis both data sets show very similar diatom-environment responses. The first CCA axis is strongly correlated with pH and calcium in both data sets indicating strong relationships between the diatom abundances and these two variables. Aluminium concentrations are negatively correlated with the first axis of the CCA in both data sets. Altitude and maximum depth are correlated with both the first and second axes of the CCA in both analyses, with TOC being negatively correlated with these two variables in both the training sets.

Forward selection of the environmental variables in the CCA of the UKAWDDS has shown that seven variables explain statistically significant amounts of the variance in the species data. These seven variables (pH, calcium, labile aluminium, altitude, potassium, TOC and maximum depth) are also shown to be important in determining the species composition of samples in the SWAP calibration set, with pH, calcium, labile aluminium and TOC/DOC being particularly related to the distributions of diatoms in both training sets.

The two training sets have been re-analysed using a common set of environmental variables to confirm the findings outlined above. CCA was used to relate the diatom abundances to the ten environmental variables included in the analysis (total aluminium, TOC, sulphate, potassium, conductivity, calcium, magnesium, pH, equivalent alkalinity and chloride). For both the data sets, only those diatom taxa that were present at 2% of the sample or greater in at least one sample were included in the analyses.

As with the comparison between the published SWAP analysis and that performed on the UKAWDDS, the new analysis confirms that the species environment relationships in the two training sets are very similar.

The ten environmental variables explain similar amounts of the total variance in the species data, with CCA axis 1 explaining 10.8% of the variance in the SWAP diatom data and 10.2% in the UKAWDDS. The second CCA axis explains 4% of the variance in the SWAP species data and 3% in the UKAWDDS. This suggests that there is a strong primary gradient in both of the training sets and a weak secondary gradient.

Many of the variables included in the analyses are closely correlated in the training sets, such as pH, calcium and equivalent alkalinity. These variables all have high variance inflation factors indicating their multi-colinearity and as such it is not appropriate to interpret the regression coefficients as these may be unstable in such circumstances. Instead, inter set correlations can be used as a guide to those variables that are correlated with the axes of the CCA, and therefore important in explaining the distributions of the diatom taxa throughout the training sets.

In both training sets pH, calcium and equivalent alkalinity are the three variables with high inter set correlations with the first CCA axis, illustrating that the acidity gradient is the most important determinand for explaining diatom abundances in the two training sets.

TOC (0.6775) and to a lesser extent total aluminium (0.438) are correlated with the second CCA axis in the SWAP training set whilst total aluminium (0.4269), chloride (-0.4092) and TOC (-0.3810) concentrations are the two variables most correlated with CCA axis 2 in the UKAWDDS. The second axis of the CCA explains very little of the species data in both

training sets but it is interesting that the secondary gradient in both training sets are also very similar.

From the comparison and analysis outlined above it is clear that the SWAP and UKAWDDS training sets exhibit very similar diatom-environment relationships and that they cover a similar range of hydrochemical gradients. This is despite the reduced geographical range encompassed in the UKAWDDS when compared to the diatom data from northern Europe covered by the SWAP calibration set.

3.5 Discussion

The first step in developing the analogue matching approach to setting targets for pre-impact conditions in acidified lakes has been to increase the number of samples in the diatom training set. The diatom training set used by Flower *et al.* (1997) contained a large number of samples from lakes in Norway and Sweden. The data set presented in this chapter is based on the data set used by Flower *et al.* (1997), but restricted to those samples from UK lakes. This basic data set has been expanded upon by incorporating new samples from a range of acidified and acid-sensitive lakes from across Scotland and Wales. This new data set, the UKAWDDS, contains 163 samples from 151 unique lakes from across the range of acid-sensitive, oligotrophic upland lakes in the UK. An analysis of this data set is presented in this chapter and analogue matching using this data set is described in Chapter 5.

Flower *et al.*'s study (1997) showed the potential of using analogue matching in the study of recovery of lakes from acidification. Their work also raises a number of methodological questions regarding the method of analogue matching. These methodological questions have already been described in Chapter 2. In this chapter, I have taken a preliminary look at the question of coverage of the appropriate environmental space by the modern training set.

The question of why the coverage of the modern training set is important in an analogue matching technique can be answered from two different points of view. From an ecological and environmental point of view the reason for encapsulating as wide a range of

ecological and environmental variation as possible in the modern training set are clear. The method of analogue matching is being used to identify those modern lakes that have similar floras and faunas to those that were found during the pre-impact phase of acidified lakes. To ensure that the matches are as analogous as possible we must ensure that the modern training set contains lakes that cover the complete range of biological and physico-chemical conditions likely to have been evident in the acid-sensitive lakes. In the UK this means creating a modern training set that is representative of conditions likely to have been found in the acid-sensitive, oligotrophic upland lakes of the UK. Apart from the considerable ecological changes that have resulted from lake acidification, these upland areas of the UK have remained much the same over the last 200 years or so. The main exception to this is the extensive coniferous afforestation in many areas of the UK. We can be confident, however, of being able to encapsulate the range of pre-acidification conditions if we sample a large enough population of lakes for the modern training set.

The second reason why the modern training set needs to be of sufficient coverage comes from the mathematical theory that underlies the calculations behind analogue matching. ter Braak (1995) has shown that analogue matching (*k*-NN) is just one of the ways of achieving regression via smoothing. Unlike more advanced methods, *k*-NN suffers more from edge effects, where bias in the inference of fossil parameters increases when the method is used to extrapolate beyond the edges of the environmental space described by the modern training set. Furthermore, in the multidimensional space of the modern training set, individual samples are actually positioned far apart along at least one dimension. By increasing the coverage of the modern training set we aim to overcome or reduce the effect of both these properties of the method. If the environmental space described by the samples is increased then it is less likely that extrapolation will be required and therefore the edge effects should be reduced. By sampling more sites, the spacing of the samples with that environmental space is reduced.

The rationale behind restricting the UKAWDDS to UK lakes is driven by the ultimate use of the data set in analogue matching. By constraining the geographic range of the data set to a single region it is hoped that biogeographical problems in higher taxa will be avoided.

3.5.1 Summary of results

The main trends in the physico-chemical data of the UKAWDDS mirror those of other data sets of similar systems (e.g. SWAP), with a strong gradient of conductivity and pH, alkalinity and calcium as the primary axis of variation. This gradient is particularly strong in the UKAWDDS and accounts for over 30% of the variance in the physico-chemical data. In this way the UKAWDDS is different to the SWAP calibration set and is a reflection of the additional sites from northwest Scotland included in the UKAWDDS training set, which experience significant sea-salt inputs (Allott *et al.* 1995).

Another important factor in the difference between the SWAP calibration set and the UKAWDDS is the inclusion of more acid-sensitive (low calcium: $< 50\ \mu$ eq l^{-1}) sites that have high pH (*ca.* pH 6.0+). These sites are likely to be potential modern analogues for acidified lakes, and the addition of these newly sampled surface waters helps to further increase the coverage of the UKAWDDS training set over that used in the original approach (Flower *et al.* 1997).

The results of the analysis of the physico-chemical data for the UKAWDDS presented in this chapter indicate that the primary aim of increasing the range of physico-chemical gradients in the data set has been achieved by the inclusion of more acid-sensitive sites, by increasing the conductivity gradient and by sampling a wider range of sites at specific pH values to capture greater variation in other important covariables (e.g. Al, Ca) at a given pH.

Constrained ordination with forward selection of the diatom data in the UKAWDDS has shown the dominant effect of pH on the distribution of diatom taxa in the training set. pH alone explains 9% of the total variance in the data set, with the seven physico-chemical variables that explain statistically significant amounts of the variance in the diatom data explaining a combined total of 21.96%. These seven variables account for over two thirds of the total variance that can be explained by the measured physico-chemical data. The other significant variables are TOC, labile aluminium, maximum depth, calcium, altitude and potassium (Table 10). The forward selection procedure in CANOCO tests the significance of the variance explained by a variable after the effects of previously tested variables have been accounted for. Therefore, the results of the forward selection procedure (Table 10

and Figure 19) indicate independent effects of each of the six variables named above on the distributions of the diatom taxa after the main effect of pH has been removed.

Similar patterns of physico-chemical response are shown in the distributions of the diatom taxa in the 83-lake subset, with pH being the most important variable. However, the relative importance of the remaining variables is different between the two data sets, with physical parameters (e.g. lake depth and altitude) being more important in the smaller data set than some of the hydrochemical variables (e.g. TOC and aluminium) are in the UKAWDDS.

3.5.2 Conclusions

The results of the analyses of the diatom data in this chapter sets clearly demonstrate the importance of pH in determining the distributions of the diatom taxa in acid, oligotrophic upland lakes. pH explains statistically significant amounts of variance (*ca.* 9% of the total variance) in both the UKAWDDS and the 83-lake subset.

The UKAWDDS encompasses a wider range of biological and physiochemical gradients than the data set originally used by Flower *et al.* (1997) because of the addition of sites from northwest Scotland, which extends the conductivity and increases the number of very acid-sensitive sites with high pH, and from other areas of upland, acid systems in Scotland and Wales, which captures greater variation in a range of variables at a given level of pH.

The wider range of sites and diatom – physicochemical responses in the UKAWDDS over those of the SWAP training set used by Flower *et al.* (1997) is particularly important for analogue matching. In the original application there were problems with the diatom – pH response masking other physicochemically-related differences between the pre-disturbance samples and the identified modern analogues. It is unlikely to be sufficient too simply add further proxies to the analogue matching procedure in an attempt to dampen the effects of the strong diatom – pH response in the training set used in the approach. The additional properties of the UKAWDDS will extend the spread of the samples in the hyper-dimensional space described by the species data and reduce the importance of acidity in identifying similar samples.

Further advantages of the UKAWDDS over the SWAP training set for analogue matching are, the reduce biogeographical range included in the UKAWDDS, and the inclusion of many more sites from areas of the UK that have received little atmospheric deposition and are considered minimally impacted with respect to acidification. The first of these additional advantages is of considerable importance because, whilst there might be limited biogeographical variation in the diatom communities between the UK and other acid sensitive lakes in Northern European, considerable biogeographical variation might be present in biological communities not used in the matching procedure (e.g., aquatic macrophytes). This would have the potential to undermine the utility of the approach for identifying appropriate analogues within the local context.

Chapter 4: Cladoceran distributions in acid-sensitive, upland lakes in the UK

4.1 Introduction

The original use of analogue matching for reference conditions in acidified surface waters in the UK was performed using the similarities of lakes based on their diatom flora (Flower *et al.* 1997). Diatoms are good indicators of the pH of the lake in which they are growing, but they are poor indicators of many other hydrochemical parameters, such as calcium concentrations. Flower *et al.* (1997) concluded that "*Identification of good modern analogue lakes can be improved by using selection criteria, other than diatoms, to pre-select sites.*", though they suggest using screening for hydrochemical or basin features to filter out inappropriate analogues. This suggestion is slightly at odds with their statement that modern analogue matching, being based on biological rather than hydrochemical criteria, is directly relevant to lake conservation and restoration objectives.

An alternative approach to the screening by physico-chemical criteria suggested by Flower *et al.* (1997) would be to base the analogue matching procedure on more biological groups than just the diatoms. The selected group or groups should be chosen such that they are indicators of those physico-chemical criteria that are important to the wider biological structure of acidified surface waters. Furthermore, the inclusion of an additional group or groups to the procedure should be restricted to those groups that are good indicators of the physico-chemical criteria that diatoms respond poorly to.

The analogue matching approach is palaeoecological in nature, and as such is restricted to those biological groups that are preserved in the sediments of lakes. A number of biological groups that fulfil this criterion were discussed previously in Section 1.3.1. An

initial decision was made to work on the sub-fossil remains of the Cladocera, and to use this group in the analogue matching procedure along with the diatoms. Cladoceran species do show a response to changes in pH, but more importantly, cladoceran community composition is determined by a complex mixture of physico-chemical and biological factors (e.g. Duigan and Kovach 1991; Duigan 1992). As such the inclusion of Cladocera in the analogue matching procedure should build in information that was absent from the diatom-only approach used by Flower *et al.* (1997).

The main aim of this chapter is to determine what the main factors are that govern the distribution of cladocera in acid sensitive fresh waters and to investigate whether the factors important in explaining cladoceran distributions are different to those factors important in explaining the distributions of diatom taxa in similar lakes.

4.2 The response of Cladocera to physico-chemical variables in upland, oligotrophic lakes in the UK

The cladoceran training set contains a series of data on 83 samples from the UKAWDDS. The 83 samples are taken from separate lakes from across the range of lakes present in the UKAWDDS. The physico-chemical data presented here are the same as that included in the diatom training set discussed in Section 3.2.1. The 83 lakes are made up of 63 from Scotland and 20 from Wales.

4.2.1 Physico-chemical properties of the 83-lake training set

As in the 163-lake diatom training set, twenty-four physico-chemical variables were measured across the 83 lake subset. Both the physical and the hydrochemical variables were available for each of the 83 lakes. pH is the geometric mean of the available data.

Summary statistics for the hydrochemical and physical data are shown in Table 14 and Table 15. The summary statistics for the 83 lake training set indicate that both the 83-lake and the 163-lake training sets cover very similar physico-chemical ranges. Like the 163-lake training set there are large ranges in pH (4.38 – 7.00, $\bar{x}$ = 5.843), calcium (11.47 – 390.7 μ eq l^{-1}, $\bar{x}$ = 85.37 μ eq l^{-1}), total aluminium (0.00 – 256.33 μ g l^{-1}, $\bar{x}$ =41.10 μ g l^{-1}), and conductivity (16.30 – 251.00 μ S cm^{-1}, $\bar{x}$ = 60.00 μ S cm^{-1}). These ranges and means compare well with those for the larger diatom training set, and indicate long acidity and conductivity gradients across both training sets.

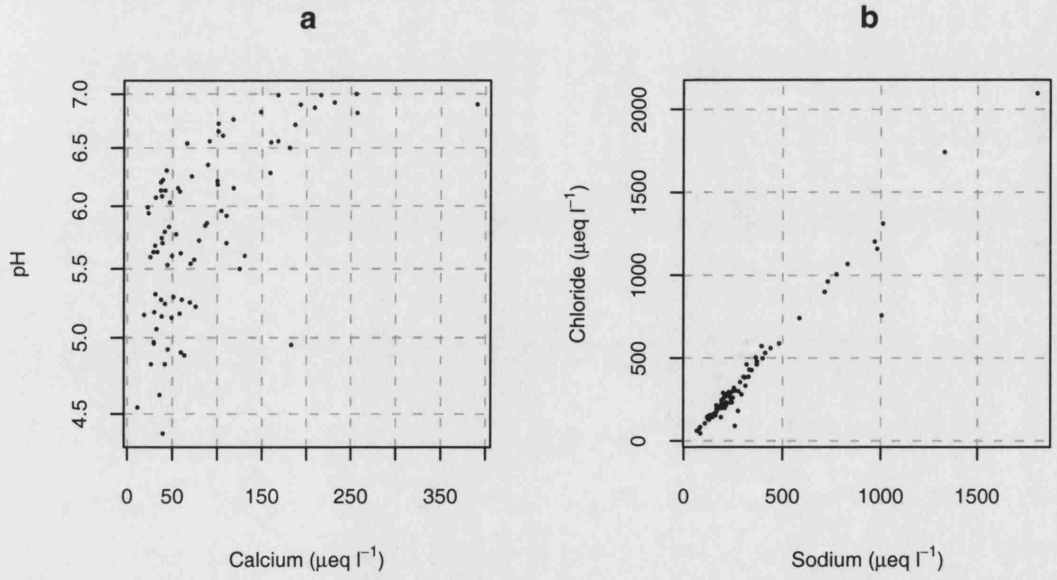


Figure 25: Scatterplots of calcium against pH (a) and sodium against chloride (b) for the 83-lake subset training set.

Table 14: Summary statistics for the hydrochemical data from the 83-lake cladoceran training set. (N = number of samples, N Miss = number of samples with missing data, Trim Mean = Trimmed mean, Std. Dev. = Standard deviation, SE Mean = Standard error of the mean, Min = minimum value, Max = maximum value, Q1 = 1st quartile, Q3 = 3rd quartile).

Variable	N	N miss	Mean	Median	Trim Mean	Std. Dev.	SE Mean	Min	Max	Q1	Q3
pH	83	0	5.8428	5.8300	5.8492	0.6708	0.0736	4.38	7.00	5.27	6.30
Conductivity (μ S cm^{-1})	83	0	60.00	46.50	54.72	44.36	4.87	16.30	251.0	33.00	67.5
Total organic carbon (mg l^{-1})	77	6	3.295	3.100	3.245	1.854	0.211	0.200	8.500	1.650	4.650
Calcium (μ eq l^{-1})	83	0	85.37	59.00	78.33	68.41	7.51	11.47	390.7	38.07	110.00
Magnesium (μ eq l^{-1})	83	0	81.39	61.00	75.64	58.91	6.47	0.0	282.00	40.17	99.00
Sodium (μ eq l^{-1})	83	0	325.7	235.0	287.6	294.9	32.4	67.0	1808.8	167.5	334.5
Potassium (μ eq l^{-1})	83	0	11.175	8.670	10.422	7.143	0.784	2.690	37.00	6.250	12.80
Sulphate (μ eq l^{-1})	83	0	77.09	64.25	72.99	44.31	4.86	21.00	233.00	45.50	90.00
Chloride (μ eq l^{-1})	83	0	380.20	260.0	332.6	366.2	40.20	45.4	2094.8	163.0	429.8
Nitrate (μ eq l^{-1})	74	9	7.63	1.55	3.33	21.36	2.48	0.00	151.00	0.42	5.00
Alkalinity (μ eq l^{-1})	52	31	60.3	28.0	45.4	87.5	12.1	-9.0	413.6	12.9	72.4
Equivalent alkalinity (μ eq l^{-1})	81	2	41.95	15.43	31.54	68.72	7.64	-17.00	394.21	2.95	51.05
Total aluminium (μ g l^{-1})	83	0	41.10	19.50	34.53	50.90	5.59	0.0	256.33	9.00	65.50
Monomeric aluminium (μ g l^{-1})	72	11	22.75	16.00	18.96	25.20	2.97	1.67	132.60	7.13	28.92
Labile aluminium (μ g l^{-1})	72	11	14.52	3.00	10.05	26.69	3.15	0.0	127.67	1.00	10.10

Table 15: Summary statistics for the physical data from the 83-lake cladoceran training set. (N = number of samples, N Miss = number of samples with missing data, Trim Mean = Trimmed mean, Std. Dev. = Standard deviation, SE Mean = Standard error of the mean, Min = minimum value, Max = maximum value, Q1 = 1st quartile, Q3 = 3rd quartile).

Variable	N	N miss	Mean	Median	Trim Mean	Std. Dev.	SE Mean	Min	Max	Q1	Q3
% Afforestation	83	0	13.59	0.00	9.84	27.18	2.98	0.00	100.00	0.00	10.00
Lake area to Depth ratio	77	6	1.870	1.060	1.512	2.252	0.257	0.150	12.130	0.580	2.500
Catchment to lake area ratio	71	12	14.82	8.33	11.39	19.14	2.27	1.50	117.80	4.75	15.33
Catchment Area (ha)	71	12	476	84	183	1551	184	2	11637	33	257
Lake Altitude (m)	82	1	291.1	236.5	273.6	215.6	23.8	10.0	1000.0	117.3	423.8
Lake Area (ha)	77	6	26.21	10.00	18.45	56.71	6.46	1.00	473.0	4.50	28.50
Maximum Altitude in Catchment (m)	66	17	568.5	544.0	560.1	339.9	41.8	47.0	1296.0	257.0	813.0
Maximum Lake Depth (m)	81	2	13.49	10.00	12.52	10.57	1.17	1.000	51.00	5.10	19.75
Net Catchment Relief (m)	65	18	278.6	210.0	261.9	250.8	31.1	4.0	935.0	74.5	444.0

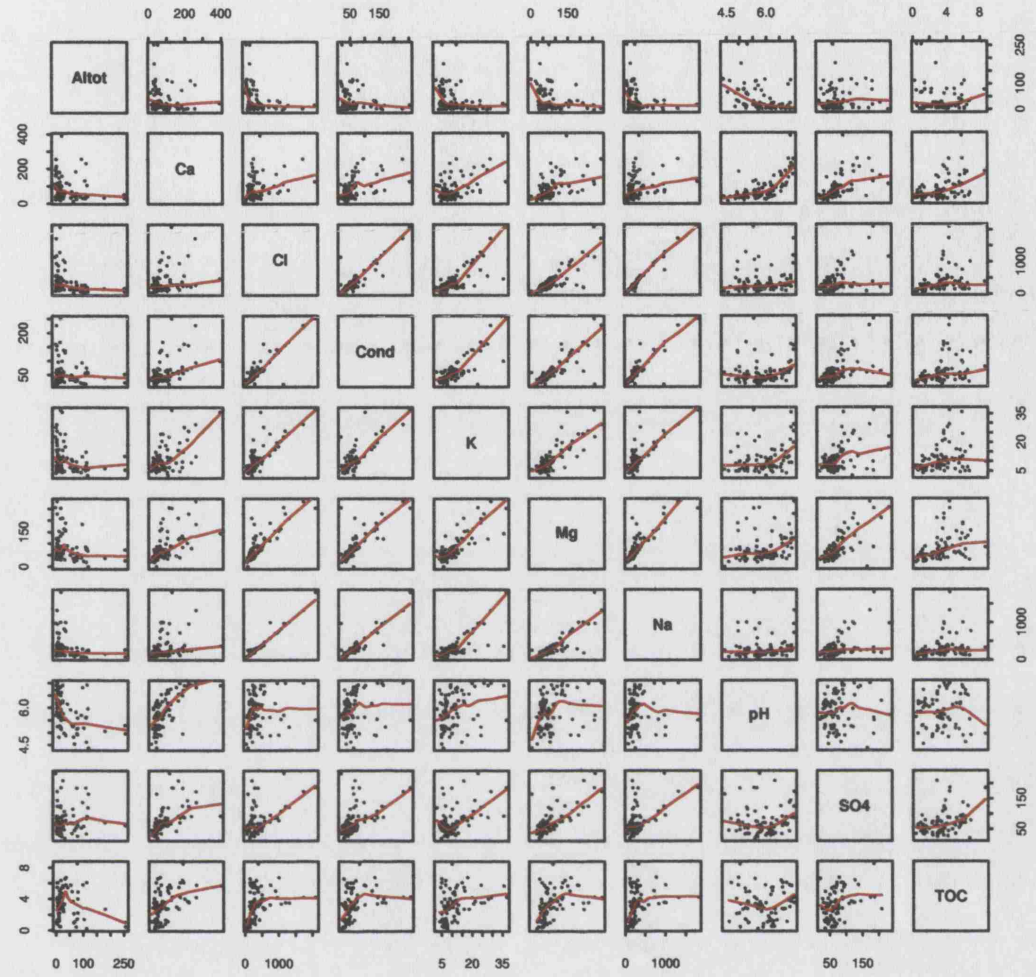


Figure 26: Scatterplot matrix of the major hydrochemical variables from the 83 cladoceran training set. Overlaid on each plot is a LOWESS scatterplot smoother (span=0.66) illustrating the correlation between each pair of variables.

Table 15 shows that, like the hydrochemical variables, the physical variables measured across the 83-lake training set are comparable to the range of values measured in the 163-sample training set. This suggests that both training sets cover a similar amount of environmental space, and that they are directly comparable in terms of reflecting the physico-chemical nature of oligotrophic, acid-sensitive surface waters in the UK.

Figure 25a shows a scatterplot of calcium against pH for the 83 hydrochemical samples in the subset training set. The plot illustrates that the 83-lake subset contains lakes from a similar range of acidities and acid-sensitivities as the larger UKAWDDS. The 83-lake subset also contains those lakes that have low (<50 μ eq l^{-1}) calcium concentrations, which are very

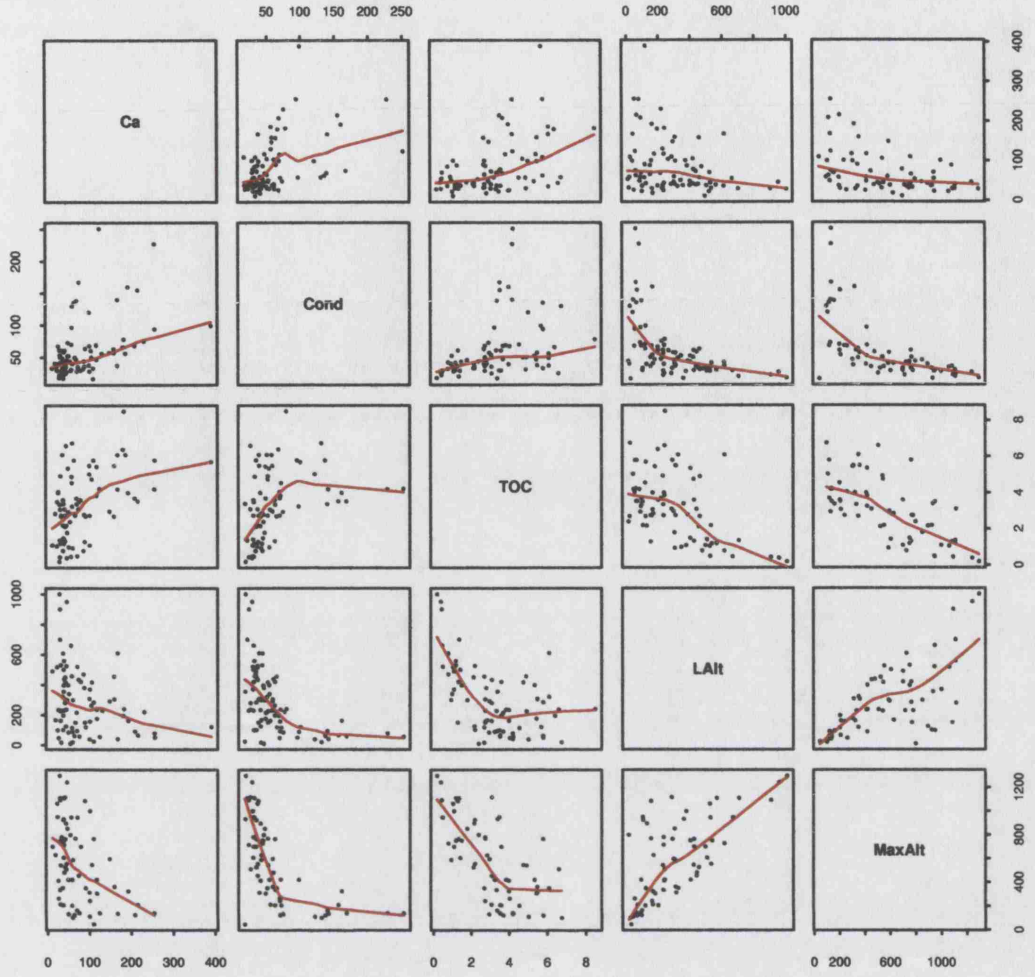


Figure 27: Scatterplot matrix of selected physico-chemical variables from the cladoceran training set. A LOWESS scatterplot smoother (span=0.66) is plotted through the data points, illustrating the main trends in the data.

sensitive to acid deposition, yet have a relatively high pH. This is an important consideration because the ultimate end-use for this training set will be analogue matching for reference conditions in acidified systems. Figure 25b shows the relationship between sodium and chloride for the 83 hydrochemical samples in the training set.

The plot illustrates that the variables are present in approximately marine proportions, suggesting that contributions are predominantly precipitation-derived. There is much less deviation from this relationship in the cladoceran training set than in the UKAWDDS, indicated by the reduced scattering of points, especially at the lower end of the gradient.

A scatterplot matrix of the hydrochemical data is shown in Figure 26. This plot shows only those variables for which the data set contains a complete, or almost complete, record.

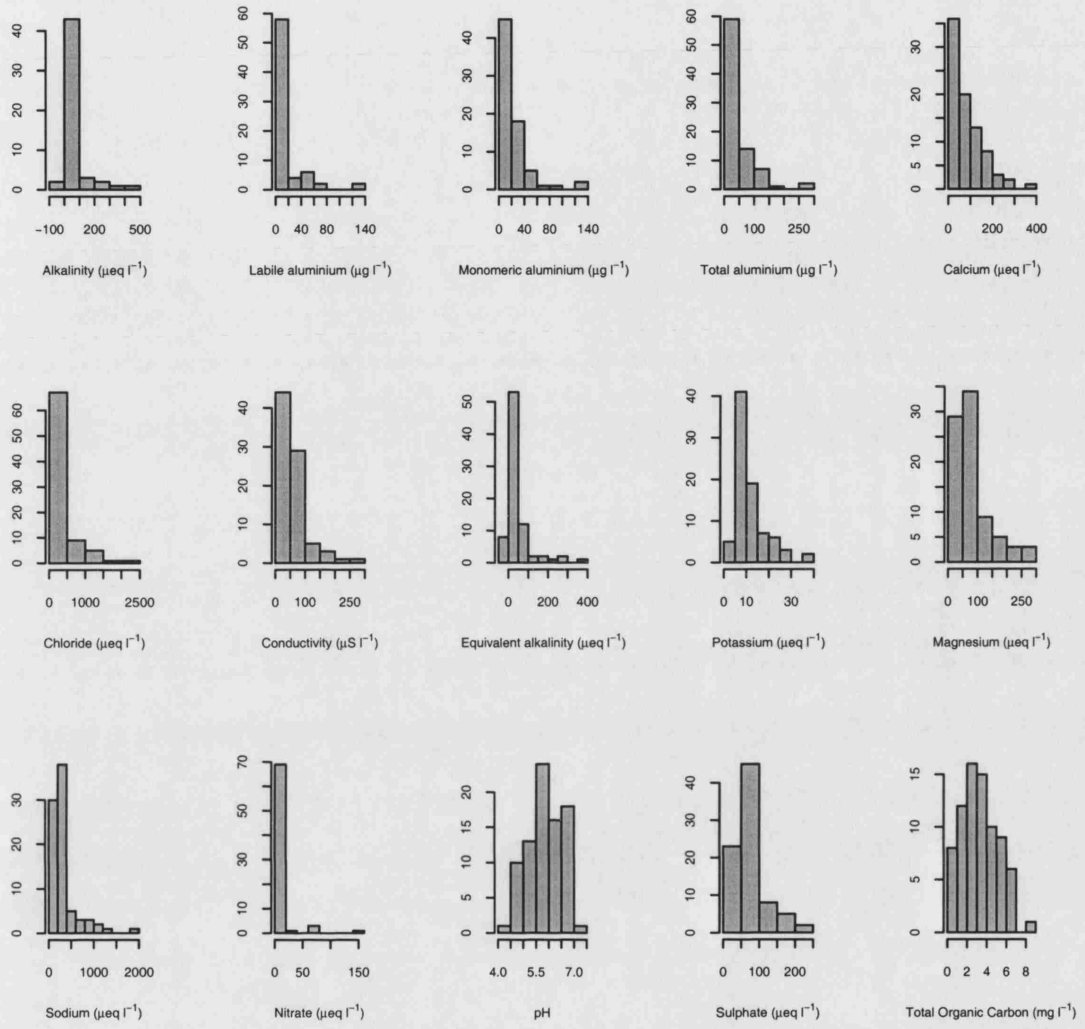


Figure 28: Histograms showing the distributions of the various hydrochemical determinands for the 83-lake training set.

A Lowess (Cleveland and Devlin 1988) smoother (span=0.66) is plotted through the data points. The relationships identified for the UKAWDDS hold true for the smaller training set also. There are strong linear relationships between conductivity and the major ions, as well as between those variables associated with acidity (pH, Ca, total aluminium, TOC and sulphate).

Figure 27 shows a scatterplot matrix of the relationships between selected physical and hydrochemical variables. The plot shows that the amount of TOC in the lake water declines with increasing lake altitude and maximum altitude in the catchment, suggesting that at higher altitudes, less carbon is transported to the lake from catchment soils. Calcium

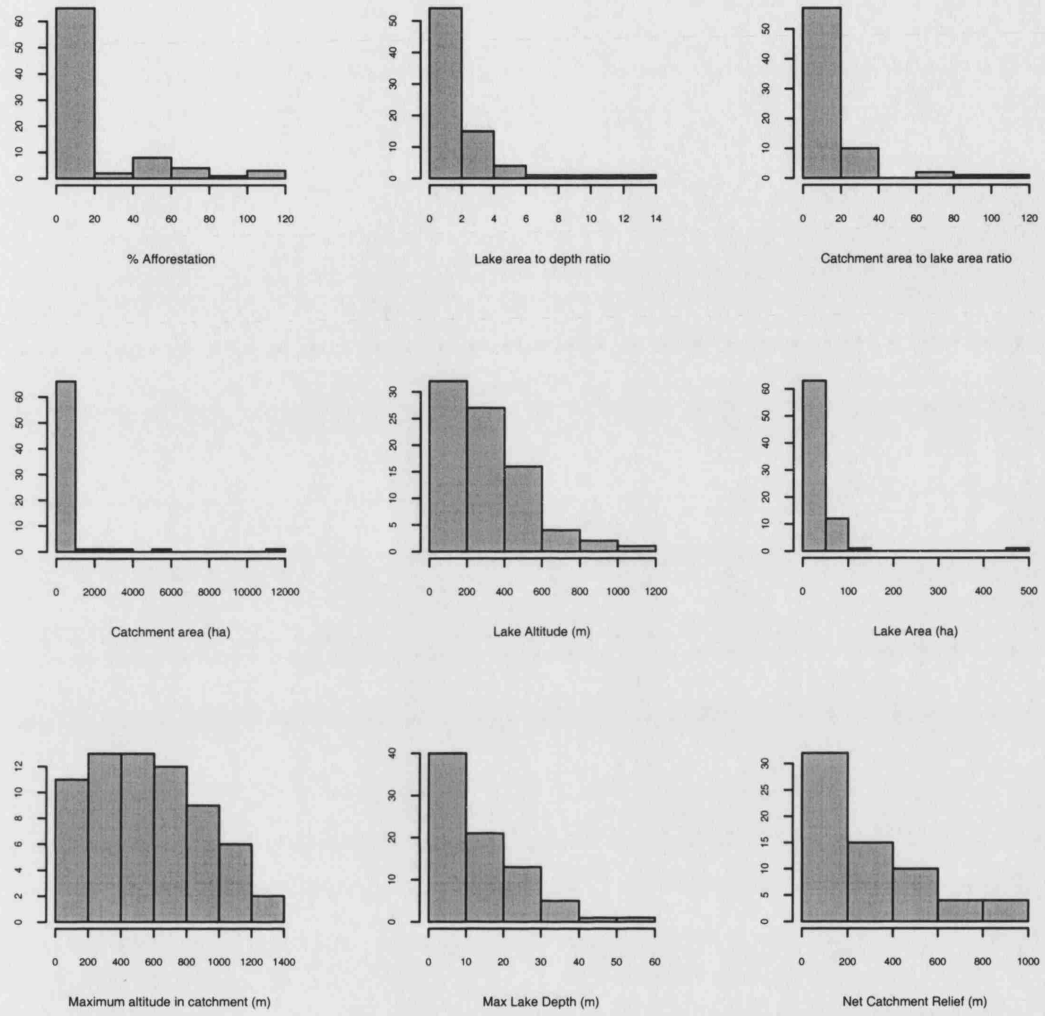


Figure 29: Histograms showing the distributions of the various hydrochemical determinands for the 83-lake training set.

concentrations also decline with increasing altitude indicating that the most sensitive lakes in the training set are also those lakes that are found at higher altitudes. Again this is probably linked to the soils in the catchments, with the catchments of higher altitude sites containing less well developed soils that are poorly buffered against acid deposition. Conductivity is also lowest in the higher altitude lakes.

Figure 28 shows histograms of the hydrochemical data for the cladoceran training set. The number of bins was chosen using Scott's rule (Scott 1979). All the variables, except pH and total organic carbon, are strongly right skewed, with long tails to the distribution and higher numbers of small values. In the subsequent quantitative analyses, with the

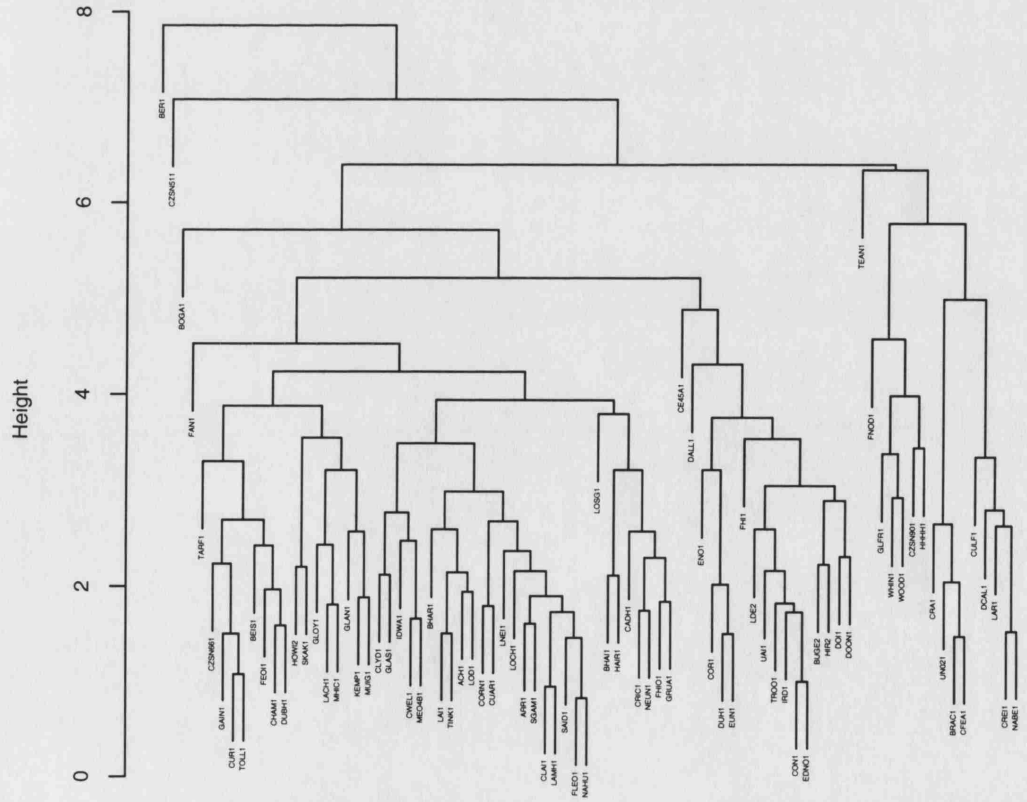
exception of pH and total organic carbon, all hydrochemical variables were $\log_{10}$ transformed prior to the analysis.

The histograms (Figure 28) show that the majority of the sites are acidic, oligotrophic and clear water lakes, with low modal calcium (<50 μeq l^{-1}), conductivity (33 – 66 μS cm^{-1}) and equivalent alkalinity (<50 μeq l^{-1}) classes.

Figure 29 shows histograms of the physical variables for the 83 lake cladoceran training set. Scott's rule (Scott 1979) was used to determine the optimal number of bins for each variable. As with the 163-sample training set, the physical variables are strongly right-skewed, though the skewness is exaggerated in the lake area and catchment area-related variables by two of lakes at the extreme high end of the gradient.

With the exception of maximum altitude in the catchment, all other physical variables were $\log_{10}$ transformed before subsequent analyses were performed, unless noted otherwise in the text. These histograms represent the properties of the data illustrated in Table 15,

Figure 30: Dendrogram showing the results of the hierarchical cluster analysis of the cladoceran data set.



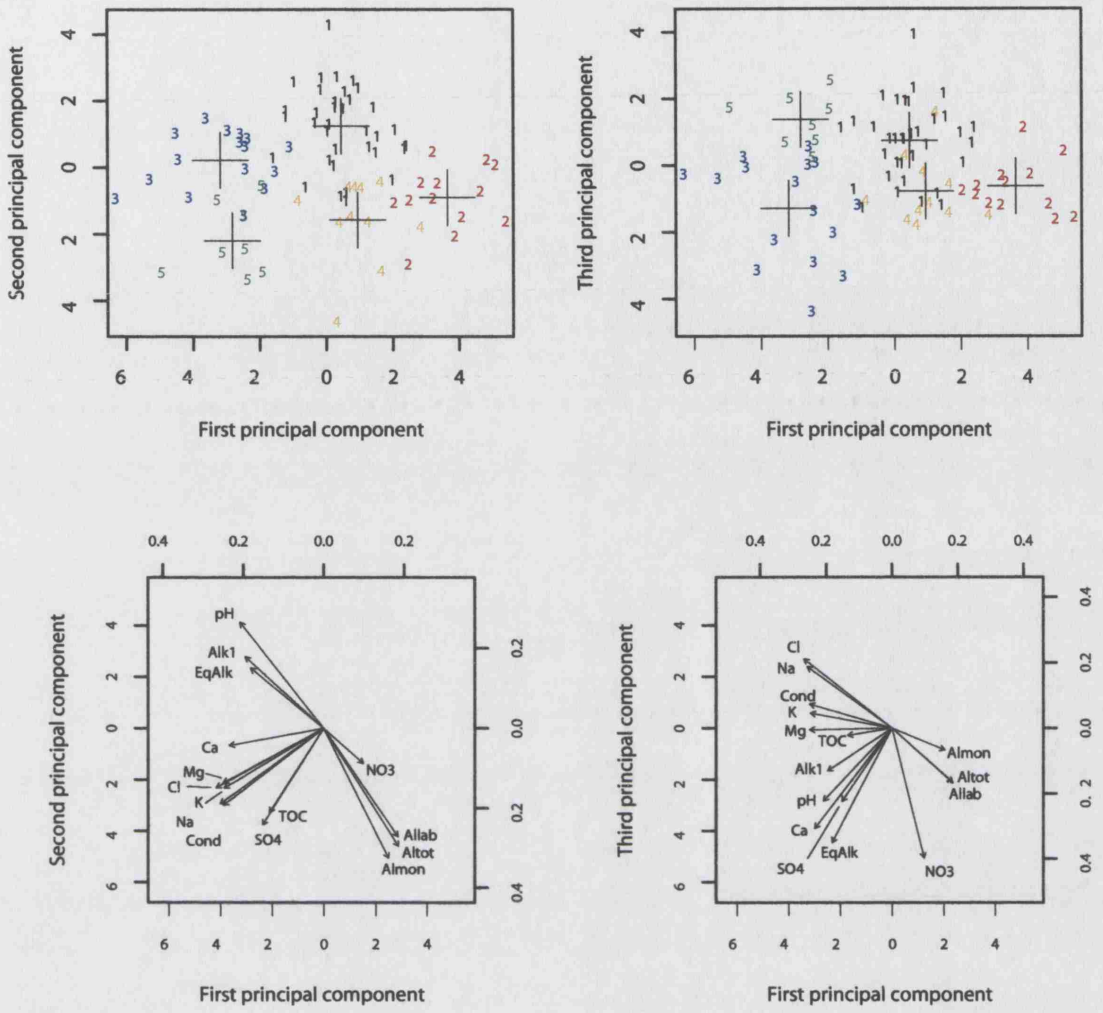


Figure 31: Biplots showing the results of the k-means cluster analysis of the hydrochemical data in the UKAWDDS. The results of the cluster analysis are projected into PCA space the cluster centroid is plotted. Below each cluster plot, biplot arrows show the importance and direction of the of the hydrochemical variables along the plotted principal components.

namely that the majority of lakes contain little or no afforestation in their catchments and are relatively small (LArea; mode = <50 ha).

The 83-lake cladoceran training set was further analysed using a cluster analysis of the hydrochemical data. *k*-means cluster analysis was used to classify the 83 lakes in the training set. The *k*-means algorithm needs starting points for the iterative algorithm, so a hierarchical-agglomerative cluster analysis (average linkage, Euclidean distance) of the standardised (zero mean, unit variance) hydrochemical data was performed first. The dendrogram (Figure 30) indicated the presence of four or five main groups of sites, with a few outliers. The dendrogram was cut automatically into five clusters. The cluster membership of each site was then used as the starting points for the *k*-means algorithm.

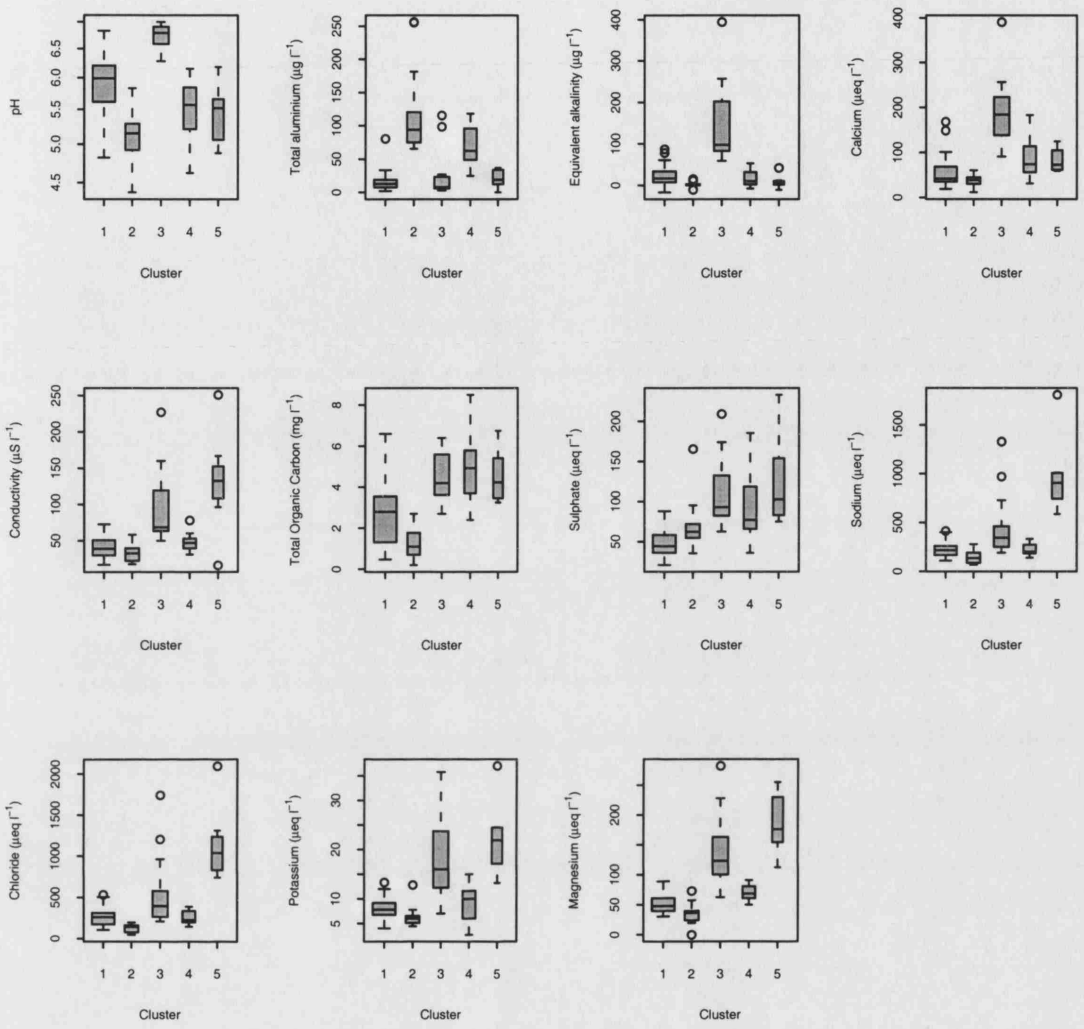


Figure 32: Boxplots of the hydrochemical variables plotted for each cluster identified by the k-means cluster analysis. The width of the boxes is proportional to the variance across the cluster of the variable.

The results of the *k*-means cluster analysis were then projected into the principal component space of the hydrochemical data. Figure 31 shows the plot of the first and second, and the first and third principal components.

The first and second principal components show a good degree of cluster separation, though there is some slight overlap between clusters. The first principal component separates the clusters in to three groups. Clusters 3 and 5 are located on the left of the plot, 1 and 4 in the centre and cluster 2 on the right. The biplot arrows show that the first principal component is a complex gradient, being a mixture of the effects of acidity and conductivity. The second principal component separates clusters 1 and 3 from the other

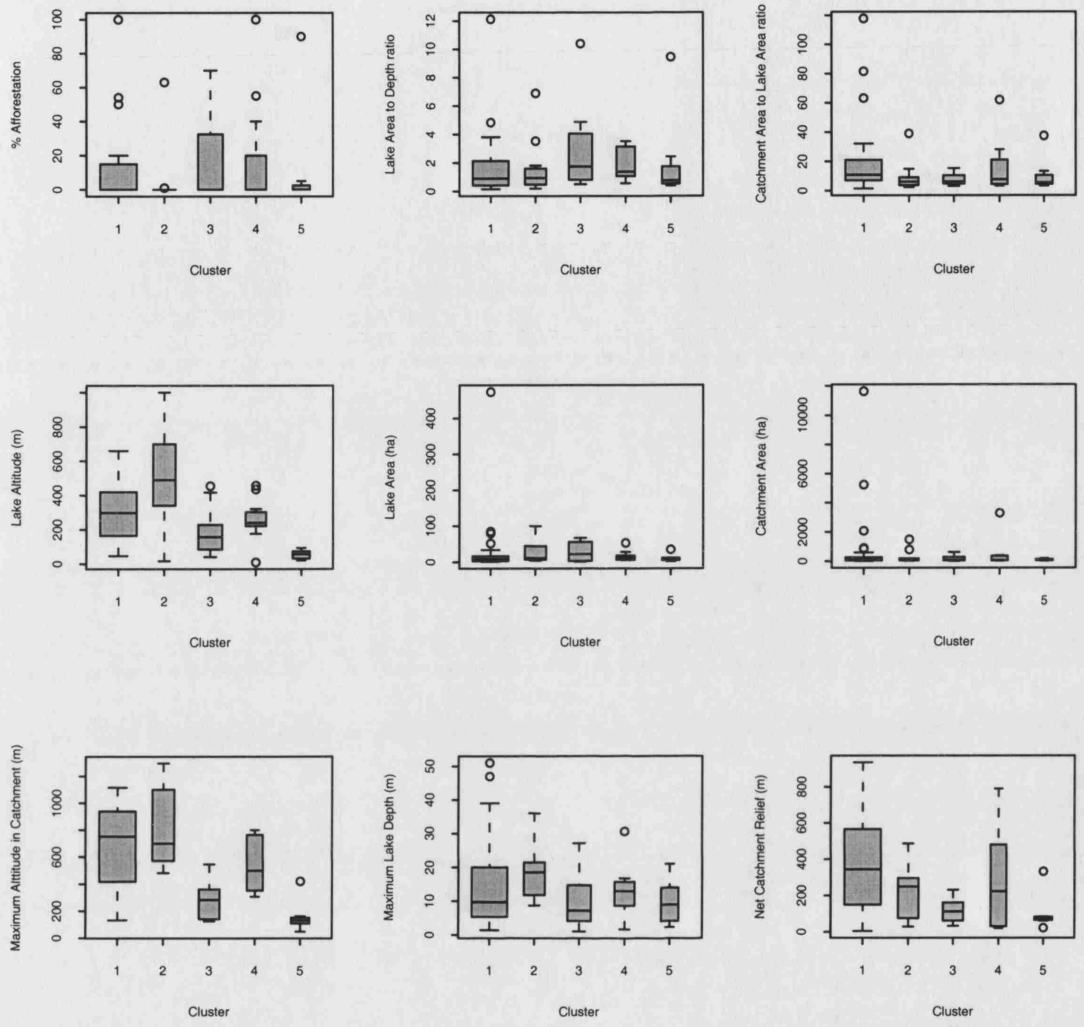


Figure 33: Boxplots of the physical variables for each of the clusters identified using k-means cluster analysis. The shaded boxes are plotted so that the width of the box is proportional to the variance of that variable for each cluster.

three groups of lakes. The biplot arrows indicate that the lakes in these two clusters are less acidic than the other three groups.

Some general characteristics of the clusters can be determined from the biplots. Clusters 1 and 4 contain lakes of approximately average conductivities, though lakes in cluster 1 are, in general, less acidic than those in cluster 4. Clusters 3 and 5 both contain lakes with above average lake water conductivity, though cluster 3 contains lakes that are less acidic than those in cluster 5. Cluster 2 contains lake of below average conductivity. The lakes in cluster 2 are also some of the most acidic in the data set.

The third principal component also shows a good degree of cluster separation. This component is strongly correlated with nitrate, equivalent alkalinity, calcium, and sodium and chloride. The lakes in cluster 3 have above average concentrations of calcium in the lake water. Clusters 1 and 2 contain lakes with the lowest calcium and equivalent alkalinity levels in the dataset. The lakes in these two clusters are poorly buffered against acid deposition. Lakes in cluster 1 are not particularly acidic, and this cluster represents those lakes in the training set that are potential analogue lakes for those lakes that are particularly acid-sensitive, and have suffered from the effects of acid deposition.

The patterns in the hydrochemical data shown in Figure 31 are also reflected in Figure 32. The boxplots show the eleven hydrochemical variables for which the data are relatively complete, plotted for each of the five clusters identified using the *k*-means cluster analysis. The plots clearly show that the lakes in cluster three are circumneutral and well-buffered against the effects of acid deposition, and this is in contrast to the four other clusters in the data. Clusters three and five contain lakes of greater ionic strength than those in the other clusters. Lakes in cluster five have much greater sodium and chloride concentrations than those in cluster three, as is the case with the overall ionic strength of these lakes.

With the exception of the lakes in cluster five, which appear to be particularly low lying, there do not appear to be any great differences in the physical parameters of the lakes among clusters (Figure 33). Lakes in cluster three also appear to be located at lower altitude than those in the remaining three clusters.

The latent structure of the physico-chemical data was further explored using principal components analysis. The variables were transformed as already discussed prior to the analysis. The hydrochemical, physical, and physico-chemical variables were subjected to separate analyses. The analysis of the hydrochemical data was conducted on eleven out of the 15 hydrochemical variables recorded in the data set (see Figure 34). Labile and monomeric aluminium, alkalinity, and nitrate were analysed as supplementary variables in the PCA to remove the effects of these variables from the analysis. As discussed above, these three variables contain many missing values (Table 14), which are interpreted as zero values in CANOCO. The missing values are not zero, rather their value is unknown, so these variables have not been allowed to take part in the ordination, but are plotted as though they did.

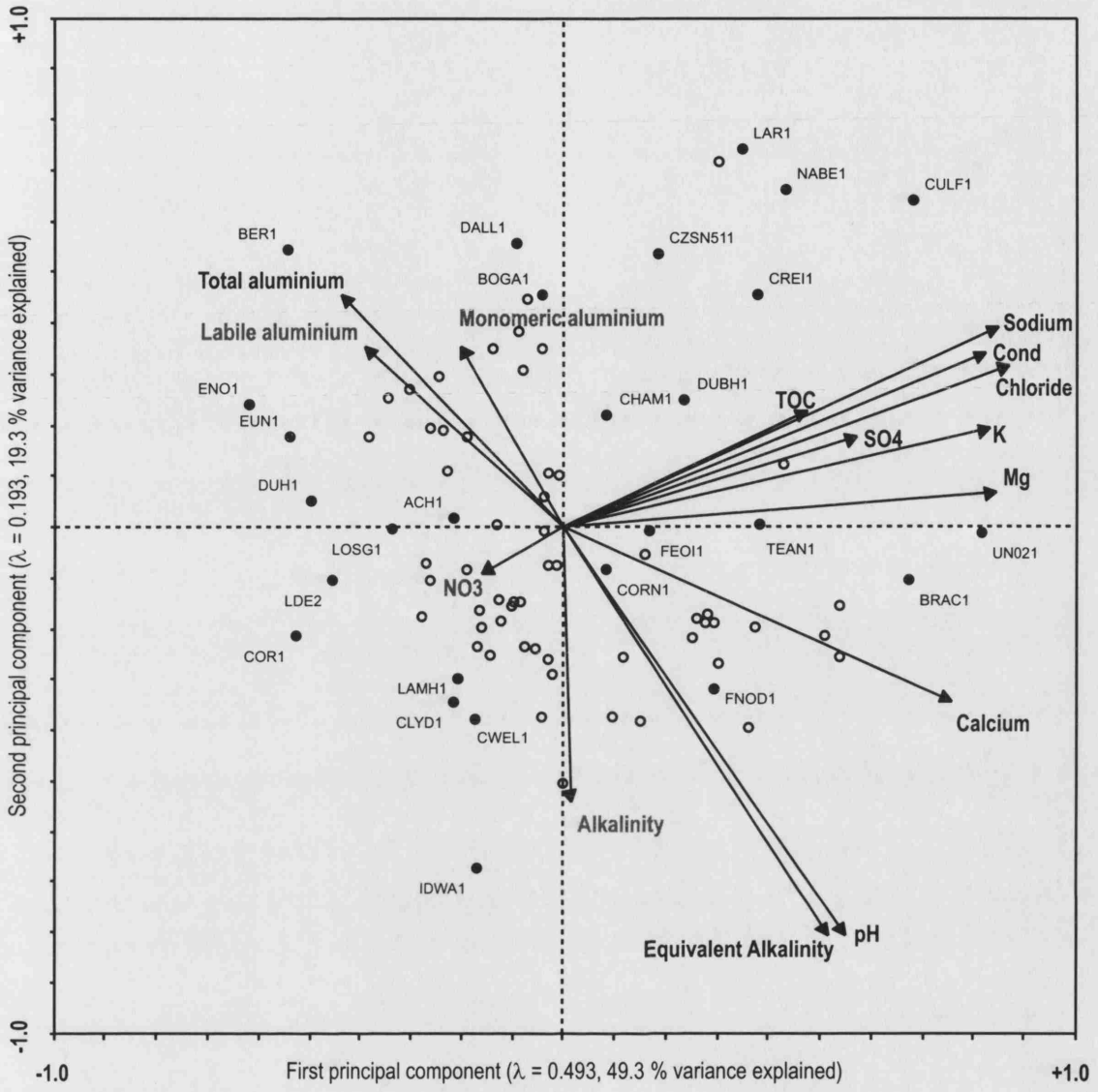


Figure 34:. Biplot showing the results of the PCA of the hydrochemical data for the 83-lake training set. The arrows labelled in grey indicate those variables analysed passively in the analysis.

The principal components analysis of the hydrochemical data indicates the presence of a single, dominant, underlying gradient. The first principal component explains 49.3% of the variance in the data set ($\lambda_1 = 0.493$, total variance = 1). Comparing the eigenvalues of the PCA to those expected under the broken stick distribution shows that the first two principal components are statistically significant, and together explain 68.5% of the variance in the hydrochemical data ($\lambda_2 = 0.193$).

The eigenvectors, or loadings, for the first principal component show it to be highly correlated with conductivity and ionic strength (e.g. chloride = 0.8674, sodium = 0.8487,

magnesium = 0.8449, potassium = 0.8290, and conductivity = 0.8227). Calcium is also highly correlated with first principal component (loading on PC_1 = 0.7519).

The second principal component (PC_2) is strongly negatively correlated with acidity (e.g. pH = −0.8056, and equivalent alkalinity = −0.8051). These two variables are also correlated with PC_1 (pH = 0.5465, and equivalent alkalinity = 0.5166), though to a lesser extent than those variables that are related to ionic strength.

Table 16: Summary results of the PCA of the hydrochemical data for the 83-lake cladoceran training set.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.493	0.193	0.110	0.079	1.000
Cumulative % variance of species data	49.3	68.5	79.5	87.4	
∑ all unconstrained λ					1.000

Table 17: Summary results of the PCA of the physical data for the 83-lake cladoceran training set.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.394	0.206	0.158	0.108	1.000
Cumulative % variance of species data	39.4	60.0	75.7	86.6	
∑ all unconstrained λ					1.000

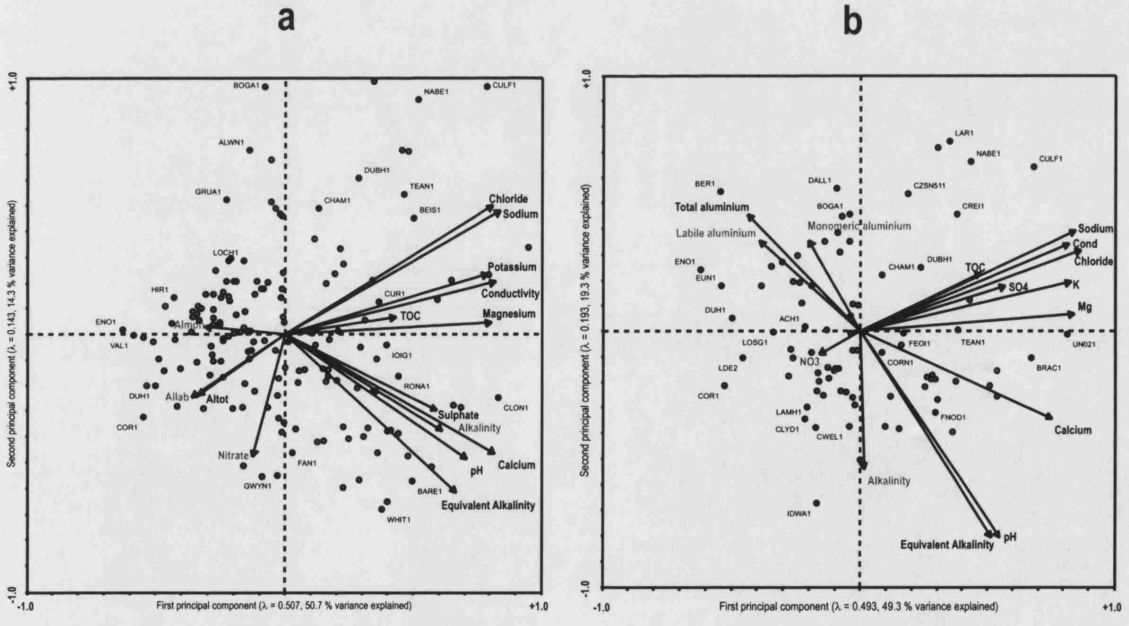


Figure 35: Comparison between the PCA of the hydrochemical of the 163-lake (a) and the 83-lake (b) data sets. The arrows labelled in red were analysed passively in the analysis.

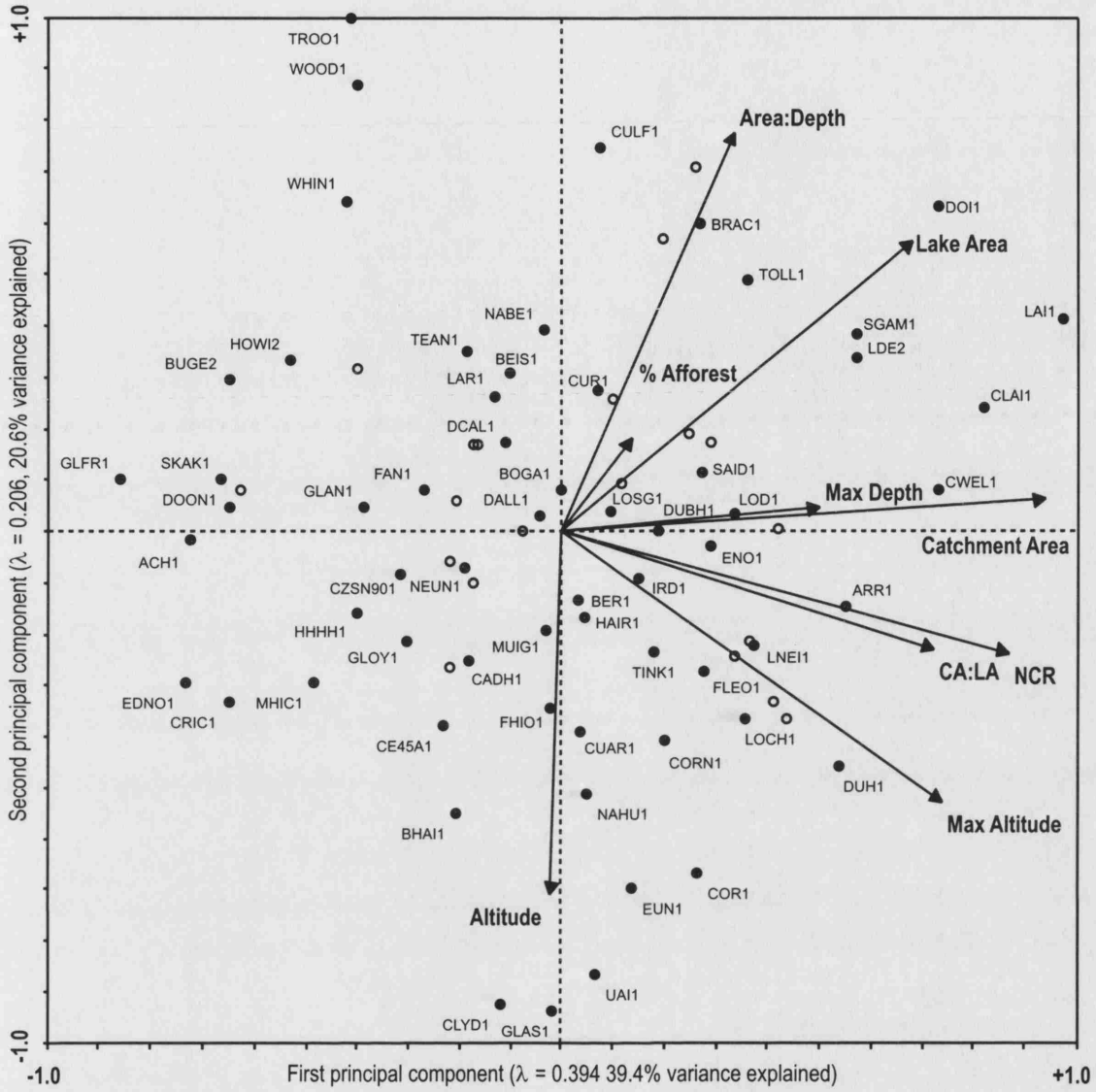


Figure 36: Principal components analysis of the combined physico-chemical data set. Arrows labelled in red indicate those variables analysed passively in the PCA.

Figure 35 shows the PCA of the hydrochemical variables from the 163-lake and for the 83-lake datasets. The biplots show a similar orientation of the eigenvectors (biplot arrows), indicating that the structure in the physico-chemical data for both datasets is similar. This suggests that the environmental space covered by both data sets is likewise similar.

The principal components analysis of the physical data indicates the presence of one main axis of variance in the 83-lake data set, which explains almost 40% of the variance in physical data. Two secondary axes, the second and third principal components, explain 20% and 15% of the variance in the physical data. The amounts of variance explained by

the first three principal components are significant when compared to the amounts of variance expected under the broken stick distribution.

The first principal component is strongly correlated with catchment area (PC_1 loading = 0.9391) and net catchment relief (PC_1 loading = 0.8696), as well as the ratio between catchment area and lake area (PC_1 loading = 0.7213), maximum altitude in the catchment (PC_1 loading = 0.6767), and lake area (PC_1 loading = 0.7377). The second principal component is correlated with the ratio between lake area and lake depth (PC_2 loading = 0.7760), and negatively correlated with lake altitude (PC_2 loading = 0.7078). Lake area and maximum altitude in the catchment also have high loadings on the second principal component of 0.5647 and −0.5308 respectively. The third principal component has high loadings for maximum lake depth (PC_3 loading = −0.5610), percentage afforestation (PC_3 loading = 0.5507), the ratio of catchment area to lake area (PC_3 loading = 0.5265), and lake altitude (PC_3 loading = −0.5003).

To look at the relative importance of the hydrochemical and the physical variables in explaining the latent structure in the physico-chemical data, PCA was used on the combined physico-chemical data set. The eigenvectors of this PCA indicate that the first principal component ($\lambda_1 = 0.3271$, $\lambda_{total} = 1.000$) is related to the ionic strength gradient (PC_1 loading for conductivity = 0.8302; chloride = 0.8250; sodium = 0.8192) identified in the analysis of the hydrochemical data (presented above), and lake altitude (PC_1 loading = −0.6199) and maximum altitude in the catchment (PC_1 loading = −0.7582). Interestingly, the first principal component is also correlated with sulphate (PC_1 loading = 0.6363).

The second principal component ($\lambda_2 = 0.188$) is correlated with physical parameters in the training set. The largest loadings on this component are catchment area (PC_2 loading = −0.8648), net catchment relief (PC_2 loading = −0.7717), catchment area to lake area ratio (PC_2 loading = −0.6652) and lake area (PC_2 loading = −0.6253). The loadings for the hydrochemical variables on the second principal component are all smaller than ±0.4, except the loading for the total aluminium concentration (PC_2 loading = 0.5502).

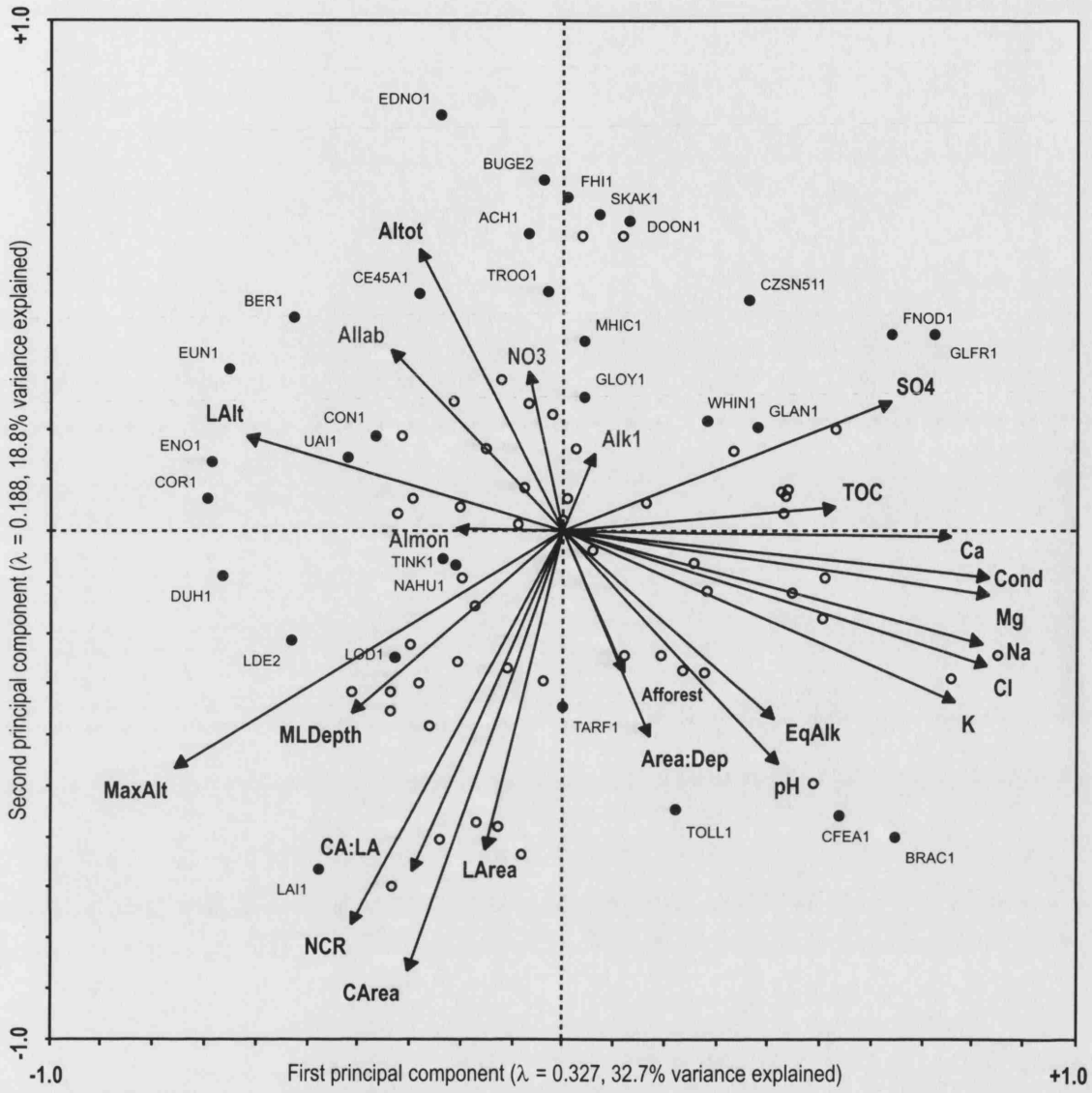


Figure 37: Principal components analysis of the 9 physical variables for the 83-lake training set.

The third principal component (λ_3 = 0.108) is an acidity gradient in the physico-chemical training set. The eigenvectors for this principal component are greatest for equivalent alkalinity (PC_3 loading = −0.7555), pH (PC_3 loading = −0.7173) and calcium (PC_3 loading = −0.5240). The eigenvectors for all other physico-chemical variables, with the exception of monomeric aluminium, are all smaller than ±0.4. The fourth principal component (λ_4 = 0.088) is strongly correlated with lake area (PC_4 loading = 0.7241) and the ration of lake area to lake depth (PC_4 loading = 0.6958).

Comparing the variances explained by the principal components with the variances expected under a broken stick distribution indicates that the four principal components

explain statistically significant amounts of the variance in the physico-chemical data set. Figure 36 shows the correlation biplot of the first and second principal components from the PCA of the physico-chemical data set. Table 18 shows the summary statistics of the PCA of the physico-chemical data for the 83-lake training set.

Table 18: Summary statistics of the Pca of the physico-chemical variables from the 83-lake training set.

Axes	**1**	**2**	**3**	**4**	**Total Variance**
Eigenvalue (λ)	0.327	0.188	0.108	0.088	1.000
Cumulative % variance of species data	32.7	51.5	62.3	71.0	
∑ all unconstrained λ					1.000

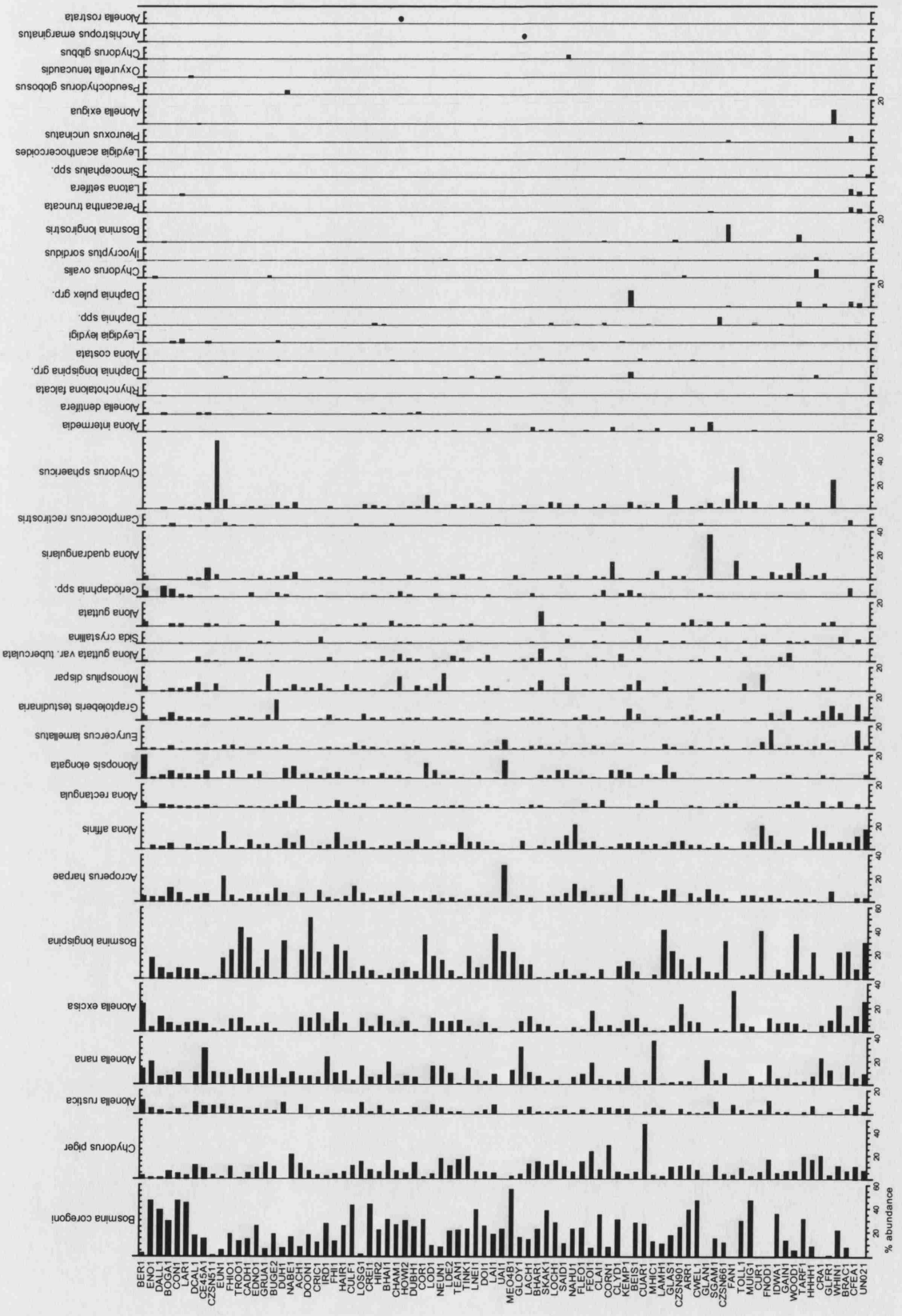


Figure 38: Percentage abundance cladoceran data for the 83-lake training set. The lakes are ordered from top to bottom by increasing pH, and the taxa are presented from left to right in increasing effective number of occurrences in the data set using Hill's N2 value.

Table 19: Summary statistics for the individual taxa in the cladoceran training set. N is the number of occurrences, N2 is Hill's N2 value and is a measure of the effective number of occurrences in the training set, Max is maximum abundance (minimum abundance is zero for all taxa), Mean and Median abundance values and SD is the standard deviation of the abundances of individual species.

Name	N	N2	Max	Mean	Median	SD
Acroperus harpae	73	43.40	30.77	5.15	4.39	4.92
Alonella dentifera	9	7.96	1.31	0.09	0.00	0.27
Alonella excisa	74	45.26	34.62	6.98	5.83	6.37
Alonella exigua	3	1.29	12.15	0.17	0.00	1.33
Alonella nana	74	47.43	36.90	8.62	6.94	7.46
Alonella rostrata	1	1.00	0.50	0.01	0.00	0.05
Alona affinis	75	43.19	20.59	4.82	3.17	4.63
Alona costata	6	4.97	1.20	0.06	0.00	0.22
Alona guttata var. *tuberculata*	37	22.74	9.84	1.16	0.00	1.88
Alona guttata	34	18.32	12.02	0.94	0.00	1.77
Alona intemedia	23	11.29	8.41	0.49	0.00	1.23
Alona quadrangularis	46	13.98	37.38	2.20	0.76	4.90
Alona rectangula	55	34.40	10.53	1.66	1.08	1.97
Alona rustica	72	50.30	11.49	3.70	3.54	2.98
Alonopsis elongatus	62	34.24	19.54	3.05	2.29	3.64
Anchistropus emarginatus	1	1.00	0.38	0.00	0.00	0.04
Bosmina coregoni	75	55.68	56.33	19.89	18.85	13.93
Bosmina longirostris	5	2.29	15.38	0.31	0.00	1.83
Bosmina longispina	76	43.40	51.11	12.94	7.93	12.36
Camptocercus rectirostris	25	13.75	4.65	0.35	0.00	0.78
Ceriodaphnia spp.	27	16.62	8.47	0.88	0.00	1.75
Chydorus gibbus	1	1.00	3.57	0.04	0.00	0.39
Chydorus ovalis	4	2.52	6.94	0.15	0.00	0.82
Chydorus piger	80	51.16	46.56	9.07	6.82	7.16
Chydorus sphaericus	51	12.10	57.14	3.20	0.76	7.75
Daphnia longispina grp.	9	5.83	4.76	0.18	0.00	0.67
Daphnia pulex grp.	6	3.41	14.29	0.36	0.00	1.75
Daphnia spp.	9	3.87	6.82	0.18	0.00	0.80
Eurycercus lamellatus	69	28.65	16.67	1.91	1.20	2.63
Graptoleberis testudinaria	58	27.73	17.01	2.19	1.27	3.09
Ilyocryptus sordidus	4	2.35	2.78	0.06	0.00	0.32
Latona setifera	3	2.10	4.65	0.09	0.00	0.58
Leydigia acanthocercoides	2	1.97	0.78	0.02	0.00	0.11
Leydigia leydigi	5	3.96	2.56	0.08	0.00	0.37
Mon dispar	53	24.91	14.50	2.23	0.90	3.41
Oxyurella tenuicaudis	1	1.00	0.45	0.01	0.00	0.05
Peracantha truncata	3	2.28	4.65	0.10	0.00	0.58
Pleuroxus uncinatus	2	1.45	4.65	0.07	0.00	0.52
Pseudochydorus globosus	1	1.00	3.33	0.04	0.00	0.36
Rhynchotalons falcata	7	6.07	1.11	0.05	0.00	0.18
Sida crystallina	34	19.22	6.02	0.72	0.00	1.31
Simocephalus spp.	2	2.00	2.33	0.06	0.00	0.35

4.2.2 Surface sediment Cladocera data

The surface sediments in the 83-lake training set were analysed for cladoceran remains following the method described in Section 2.2.2 above.

Figure 38 shows the cladoceran percentage abundance data for the 83 lakes in the training set. The lakes are ordered from top to bottom in order of increasing lake water pH, whilst the species are ordered from left to right according to the effective number of occurrences in the training set samples, species to the left of the diagram having a higher effective number of occurrences than those species on the right. At this level the diagram is difficult to interpret in detail, but one factor that stands out is that at higher values of pH (bottom of the diagram) a more diverse cladoceran community is present in the surface sediments of lakes than in the acidic lakes in the training set (top of the diagram). Consequently, the proportion of the total assemblage represented by *Bosmina* species tends to be lower in the circumneutral lakes than in the acidic lakes.

Table 19 above, shows summary statistics for the cladoceran percentage abundance data. It is interesting to note that none of the taxa are found consistently throughout the training set, *Chydorus piger* being present in the greatest number of samples (80). 21 of the species of Cladocera are found in fewer than 10 of the samples in the training set.

Those taxa that were less than 2% abundant in any one sample were deleted prior to the analysis described below. After removal of the rare taxa from the Cladocera data 47 taxa remained. Percentage species composition was not recalculated following the removal of the rare taxa.

In contrast to the large amounts of variance (inertia) in the diatom data (Section 3.3 above), the cladoceran data was of low variance (total inertia = 0.932), and exhibited short responses of the cladoceran taxa to the latent environmental variables extracted by DCA. This suggests that the cladoceran distributions in samples throughout the training set are quite similar to one another and that species turnover and replacement along underlying environmental gradients is low. It is surprising that such little variation is apparent in the Cladocera data given the wealth of evidence to suggest that pH is a particularly important variable in determining Cladocera survival (e.g. Potts and Fryer 1979; Fryer 1980; Krause-

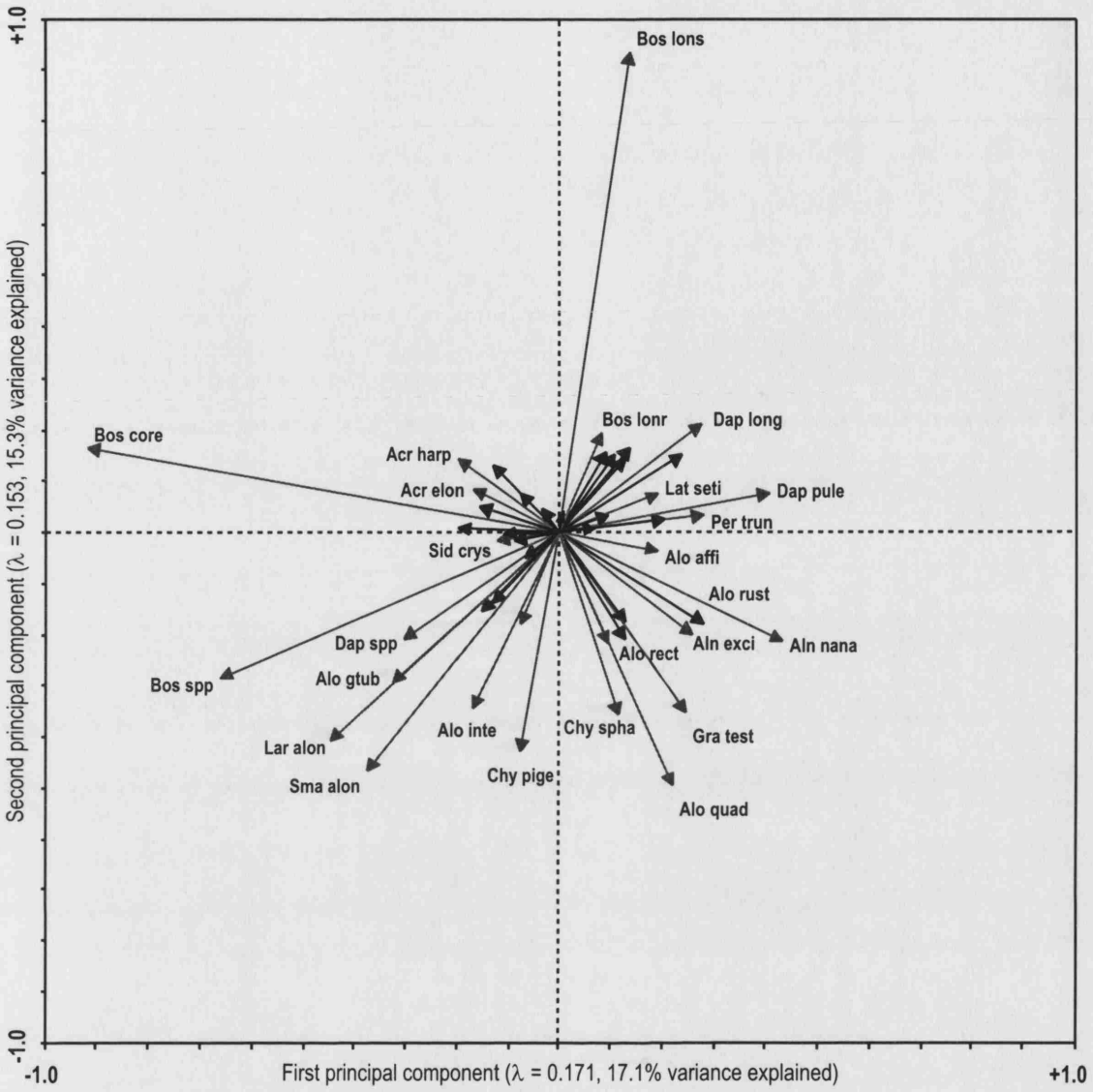


Figure 39: Principal components correlation biplot of the PCA of the cladoceran data for the 83-lake training set. Not all eigenvectors are labelled; the red arrows are those which are labelled with species codes. See appendix B for explanation of these species codes.

Dellin and Steinberg 1984; Arzet, Krause-Dellin, and Steinberg 1986; Locke 1992; Schartau, Walseng, and Snucins 2001; Walseng and Schartau 2001)

Principal components analysis of the covariance matrix of the species data was used to investigate the underlying structure in the cladoceran training set, as the species response model assumed in PCA is linear. The amounts of variance explained by each of the first four principal components of the cladoceran data are significant when compared to the amounts of variance expected under the broken stick distribution. The first two principal components explain similar amounts of the variance in the cladoceran species information, indicating the presence of two strong gradients in the data ($\lambda_1 = 0.171$, $\lambda_2 = 0.153$, $\lambda_{total} =$

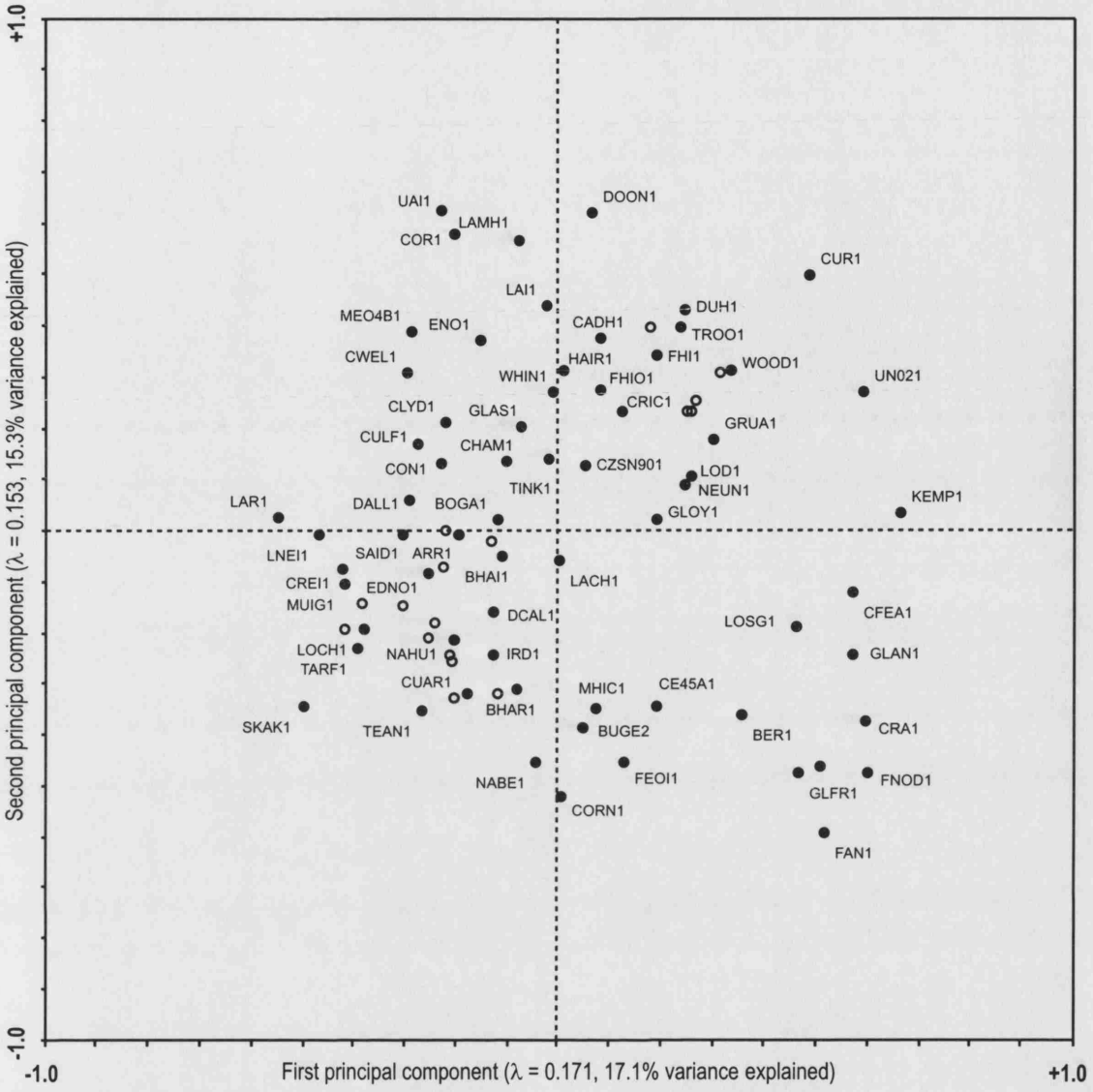


Figure 40: Principal components biplot showing the sample scores of the PCA of the cladoceran data for the 83-lake training set. Not all samples are labelled. See appendix A for sample codes.

1.000). The four axes of the PCA together explain 48.1% of the variance in the cladoceran data.

Table 20: Summary statistics of the PCA of the 83-lake cladoceran data set.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.171	0.153	0.083	0.073	1.000
Cumulative % variance of species data	17.1	32.4	40.8	48.1	
∑ all unconstrained λ					1.000

Table 21: Summary statistics of the RDA of the 83-lake cladoceran data set and the 9 physical variables.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.072	0.034	0.021	0.008	1.000
Species – Environment correlation	0.691	0.629	0.552	0.453	
Cumulative % variance of species data	7.2	10.6	12.7	13.5	
Cumulative % variance of species environment correlation	52.9	78.1	93.7	100.0	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.135

Table 22: Summary statistics of the RDA of the 83-lake cladoceran data set and the 15 hydrochemical variables.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.083	0.042	0.033	0.025	1.000
Species – Environment correlation	0.742	0.715	0.686	0.519	
Cumulative % variance of species data	8.3	12.6	15.9	18.4	
Cumulative % variance of species environment correlation	32.7	49.3	62.3	72.2	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.255

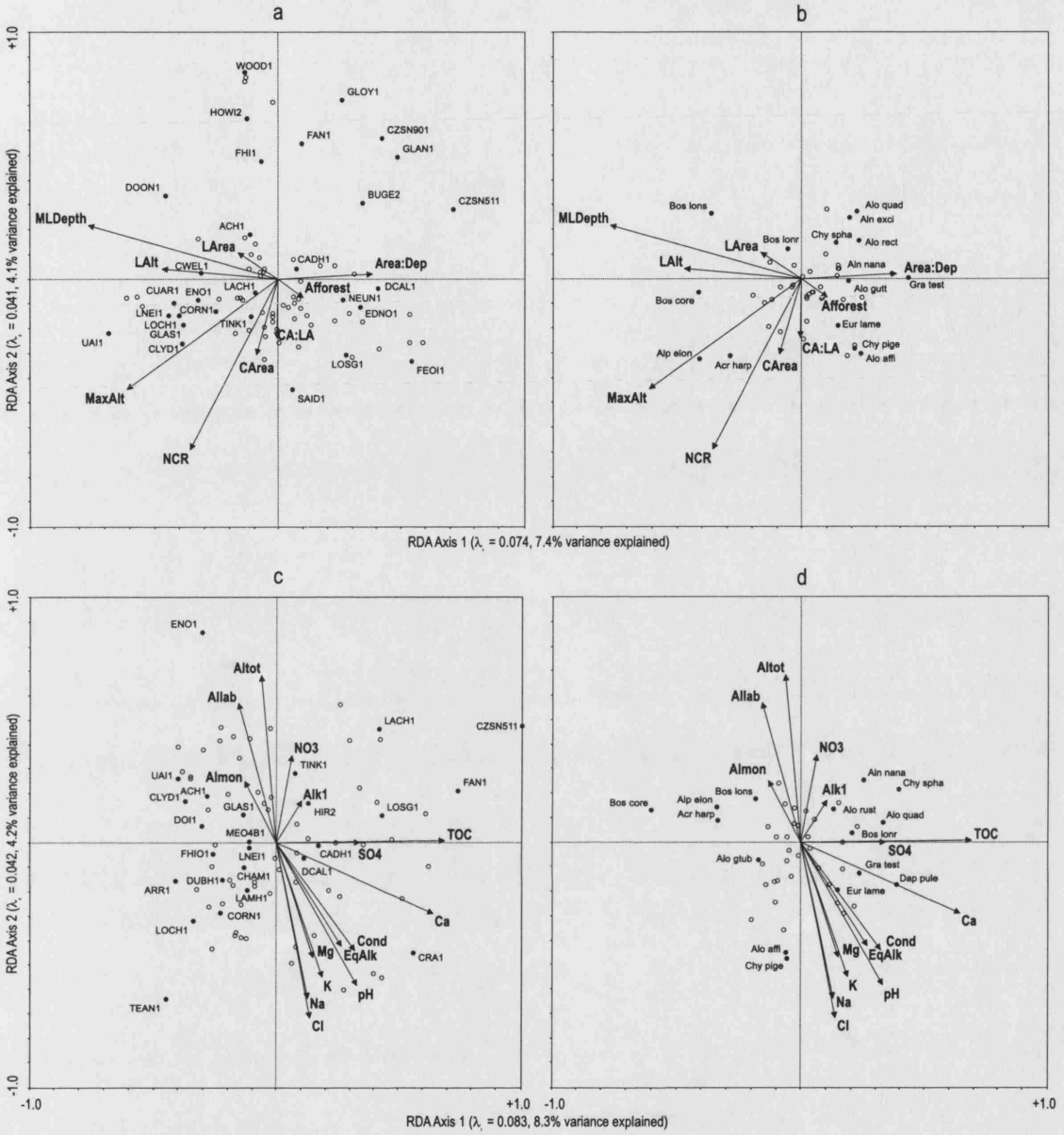


Figure 41 a-d: Redundancy analysis biplots of the cladoceran dataset. a) Site and environment biplot, and b) species and environment biplot, of an RDA of the Cladocera and the physical data, c) site and environment biplot, and d) species and environment biplot, of an RDA of the Cladocera data and the hydrochemical data.

The relationship between the Cladocera and the measured environmental variables was examined using redundancy analysis (RDA). The relationships between the physical and the hydrochemical variables and the Cladoceran data were initially analysed separately and then in a combined analysis with the 24-physico-chemical variables. The results of these gradient analyses are shown in Table 21, Figure 41a, and Figure 41b, Table 22, Figure 41c, and Figure 41d, and Table 23 respectively.

The 9-physical variables explain 19.3% of the total variance in the cladoceran data, whilst the 15-hydrochemical variables explain 20.6% of the total variance. When the physical and hydrochemical data sets are combined, the amount of variance explained increases to 40.5%. This suggests that the variance explained by the physical and the hydrochemical variables is largely unique to the individual sets of variables. This observation will be discussed in more detail later using the methods of variance partitioning and partial constrained ordination techniques.

Table 23: Summary statistics of the RDA of the 83-lake cladoceran data set and the 24 physico-chemical variables.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.113	0.066	0.043	0.035	1.000
Species – Environment correlation	0.838	0.774	0.747	0.723	
Cumulative % variance of species data	11.3	17.9	22.2	25.7	
Cumulative % variance of species environment correlation	27.8	44.2	54.7	63.4	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.405

Regression coefficients, also known as canonical coefficients, are the best weights for the respective variables along each axis. Examining the regression coefficients of the RDA and their associated t-values gives an approximate indication as to which of the environmental variables are important in determining the distribution of the Cladocera.

The regression coefficients (b) for the first axis of the RDA with the 9-physical variables show that max depth (b = -0.9183) and catchment area (b = 0.821) are important variables in explaining variance along the first axis of the RDA. Statistically significant (p=0.01) regression coefficients for the second axis of the RDA show that the ratio of catchment area to lake area (b = 1.2247), net catchment relief (b = -1.1770) and catchment area (b = -0.9805) are important in explaining variation in the Cladocera along this gradient. The variables mentioned above also have statistically significant regression coefficients for RDA axes 3 and 4, with maximum altitude also being important on axis 3.

In situations where the environmental variables are collinear (i.e. the variance inflation factor [VIF] for an environmental variable is greater than 20), the regression/canonical coefficients may become unstable (ter Braak 1986). In such cases, the inter-set correlations of the ordination may be a better indicator of the relationship between the extracted axes and the environmental variables. The inter-set correlations are the correlation coefficients between the environmental variables and the sample scores that are derived from the species scores (ter Braak and Smilauer 2002).

Table 24 shows the inter-set correlations of the RDA of the 83-lake data set and the 9 physical variables. Lake depth shows a high correlation with the first axis, but in contrast with the canonical coefficient, catchment area appears to be uncorrelated with this axis. Of the variables with significant canonical coefficients for the second RDA axis only NCR shows a strong correlation with this gradient.

Table 24: Inter-set correlations of environmental variables with axes

Variable	Axis 1	Axis 2	Axis 3	Axis 4
% Afforestation [Afforest]	0.0730	-0.0520	-0.3147	-0.1022
Area to Depth Ratio [Area:Dep]	0.2761	0.0127	-0.2709	-0.1472
Catchment to Lake Area ration [CA:LA]	-0.0052	-0.1617	-0.1200	0.0445
Catchment Area [CArea] (ha)	-0.0640	-0.2051	-0.1869	-0.1836
Lake Altitude [LAlt] (m)	-0.3354	0.0286	0.3491	0.0378
Lake Area [LArea] (ha)	-0.1180	0.0697	-0.3107	-0.2853
Maximum Altitude in Catchment [MaxAlt] (m)	-0.4345	-0.2989	0.0963	-0.0752
Max Lake Depth [MLDepth] (m)	-0.5434	0.1445	-0.2228	-0.1900
Net Catchment Relief [NCR] (m)	-0.2526	-0.4615	-0.1914	-0.0330

Figure 41a and b show the distribution of the sites and the species respectively with respect to the 9 measured physical parameters. Long biplot arrows indicate large loadings on the first RDA axis for max lake depth, on the second axis for NCR and on both first and second RDA axes for MaxAlt.

A forward selection step was performed to determine the minimally adequate model to describe the distribution of the cladoceran species constrained by the physical parameters. Table 25 shows the results of this selection procedure, indicating that 4 of the nine physical variables explain statistically significant amounts of additional variance in the cladoceran data, as assessed by Monte Carlo permutation tests (999 permutations) and a Bonferroni type correction to the p-values required for significance. Using these four physical parameters in an RDA of the cladoceran data indicates that 13.5 % of the variance in the cladoceran data is explained by the minimally adequate model. The remaining five variables explain only 5.8% of the variance in the species data, and the explained variance is located along RDA axes 3 and 4 which explain very little of the cladoceran species distributions.

Table 25 Results of the forward selection of the 9 physical variables. Only those variables that explain statistically significant additional amounts of variance in the Cladocera data se are shown. Statistical significance assessed at a=0.05 using 999 Monte Carlo permutations tests. P-value is the significance of the F-value assessed by permutation tests. P-required is a Bonferroni corrected P-value. λ = Eigenvalue for the environmental variable and is a measure of the (extra) variance explained. % is the extra variance explained expressed as a percentage.

Variable	F	P-value	P-required	λ	%
Max Lake Depth (MLDepth)	4.420	0.001	0.05	0.052	5.2
Net Catchment Relief (NCR)	2.651	0.003	0.025	0.030	3.0
Maximum Altitude (MaxAlt)	2.571	0.003	0.017	0.029	2.9
Catchment Area (CArea)	2.175	0.01	0.0125	0.024	2.4

The analysis procedure performed above for the physical variables was repeated for the 15 hydrochemical variables, to determine how well the distributions of the cladoceran species is explained by water chemistry. The results of this analysis are shown in Table 22 and Figure 41c and d, and indicate that 25.5% of the variance in the cladoceran data is explained by variations in the hydrochemistry of the 83-lakes in the training set.

The first axis of this ordination is correlated with TOC, a measure of both water clarity and organic acid content of the lake water. High TOC values occur in lakes whose catchments drain peaty soils. Ca is well correlated with both the first axis of the RDA and the second, as demonstrated by the angles between the biplot arrow for Ca and the respective axes in Figure 41c and d. The second axis of this RDA is a complex acidity and nutrient gradient, with the most acid lakes located towards the top of Figure 41c, which also include the most nutrient poor lakes. These lakes are characterised by high proportions of *B. coregoni*, *B. longispina*, *A. nana* in the surface sediment sample, taxa known to be common in oligotrophic, acid lakes. Those lakes with higher pH are located towards the bottom of Figure 41c, and whose surface sediment samples are characterised by species such as *A. affinis* and *E. lamellatus*.

The relationship between the important variables and the ordination axis is also reflected in the inter-set correlations, shown in Table 26, with high correlations being seen for TOC and Ca on the first axis, and Na, Cl, pH and Altot on the second. A few of the 15-hydrochemical variables have high VIFs suggesting that the canonical coefficients may not be stable due to multicolinearity among the environmental variables – a significant coefficient is likely the result of this multicolinearity and is therefore, not a good indicator of the significance of those variables (ter Braak and Smilauer 2002). As a result, subsequent analysis of these relationships is restricted to the inter-set correlations.

Forward selection (999 Monte Carlo permutation tests, Bonferroni-type corrected p-values) of the environmental variables suggests that just three of the fifteen hydrochemical variables explain statistically significant ($\alpha = 0.01$) amounts of extra variance (Table 27).

Table 26: Inter-set correlations for the 15 hydrochemical variables in the RDA of the cladoceran data.

Variable	Axis 1	Axis 2	Axis 3	Axis 4
Alkalinity [Alk1]	0.0771	0.1203	-0.3586	0.1241
Labile aluminium [Allab]	-0.1172	0.4075	0.1394	-0.1322
Monomeric Aluminium [Almon]	-0.0979	0.1802	0.2825	-0.1548
Total aluminium [Altot]	-0.0476	0.4895	0.681	-0.781
Calcium [Ca]	0.4778	-0.2071	-0.3296	0.0595
Chloride [Cl]	0.1024	-0.5149	0.0653	-0.0351
Conductivity [Cond]	0.2410	-0.3172	-0.0149	-0.1366
Equivalent Alkalinity[EqAlk]	0.2002	-0.3042	-0.4787	0.0248
Potassium [K]	0.1399	-0.3962	-0.1439	-0.2531
Magnesium [Mg]	0.1131	-0.3375	-0.2763	0.0269
Sodium [Na]	0.0958	-0.4590	0.0522	-0.0732
Nitrate [NO3]	0.0464	0.2560	-0.3060	-0.1340
pH	0.2447	-0.4193	-0.3916	0.0351
Sulphate [SO4]	0.2568	-0.0032	-0.2200	0.0793
Total organic carbon [TOC]	0.5144	0.0036	0.1044	-0.0451

Table 27 Results of the forward selection of the 15 hydrochemical variables. Only those variables that explain statistically significant additional amounts of variance in the Cladocera data set are shown. Statistical significance assessed at α=0.05 using 999 Monte Carlo permutations tests. P-value is the significance of the F-value assessed by permutation tests. P-required is a Bonferroni corrected P-value. λ = Eigenvalue for the environmental variable and is a measure of the (extra) variance explained. % is the extra variance explained expressed as a percentage.

Variable	F	P-value	P-required	λ	%
Calcium (Ca)	4.023	0.0010	0.05	0.047	4.7
Total Organic Carbon (TOC)	2.582	0.0030	0.025	0.030	3.0
Chloride (Cl)	2.879	0.0020	0.017	0.032	3.2

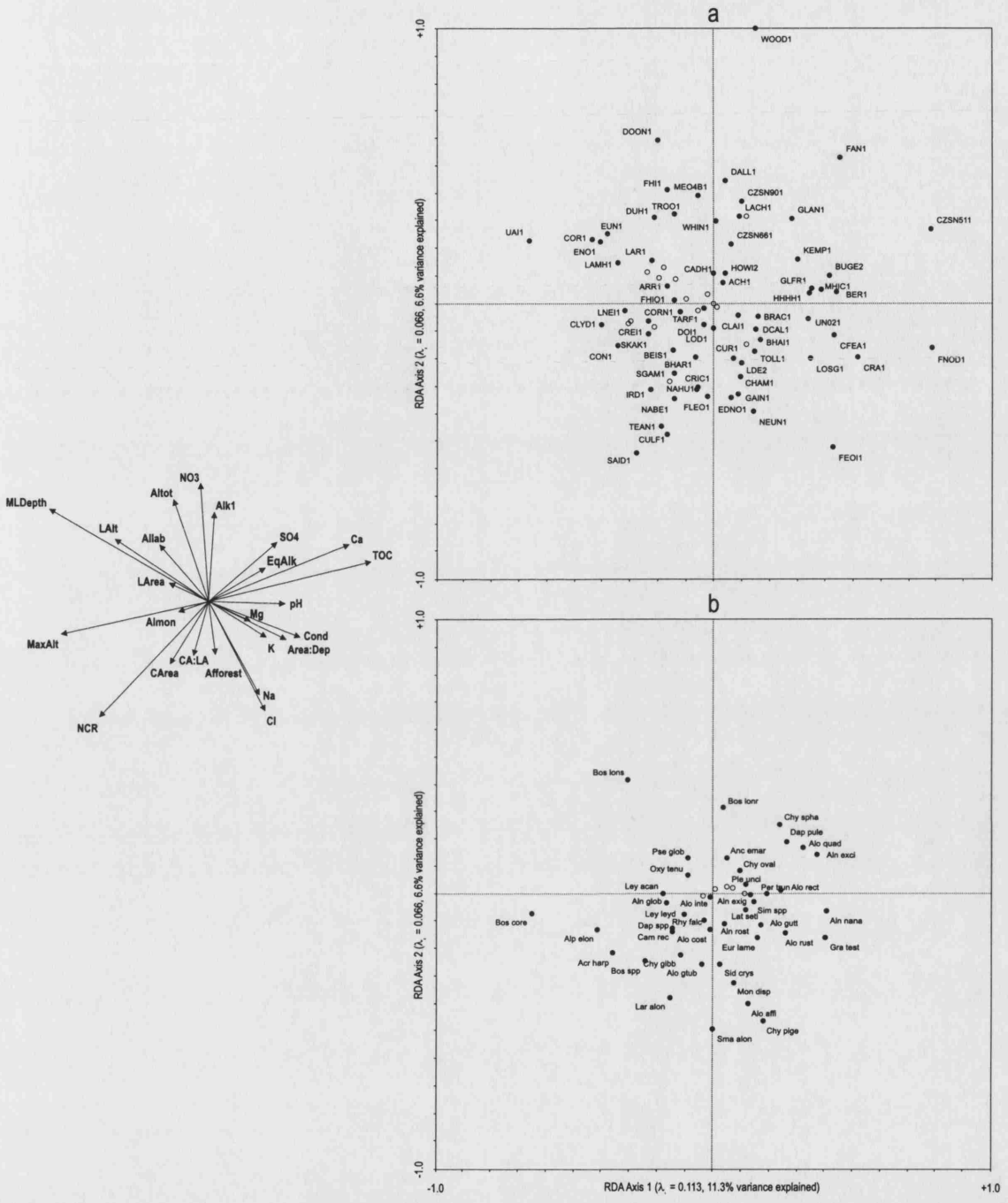


Figure 42: Results of the RDA of the 24 physico-chemical variables and the cladoceran data for the 83-lake training set. a) site environment biplot and b) species environment biplot. The biplot arrows have been removed to aid interpretation of the biplots, and are shown on the left of the diagram. See appendix A, B and C for explanations of the species, site and physico-chemical variable codes respectively.

The relative importance of the physical and the hydrochemical variables in explaining the distributions of the Cladocera were investigated using RDA of the species data constrained by the 24 physico-chemical variables. From the results of this analysis (Figure 42 & Table 23) we can see that 40.5% of the variance in the cladoceran data is explained by the 24 physico-chemical parameters; the first two RDA axes explaining almost 18% of the variance. The total variance explained by the 24 variables (40.5%) is slightly greater than the total variance explained by the two sets of variables when analysed individually (13.5% [physical] + 25.5% [hydrochemical] = 39.8%). The extra variance explained probably relates to interactions between the physical and the hydrochemical parameters that were not accounted for when the two sets were analysed separately.

The results of this ordination indicate that the parameters that were important in explaining the cladoceran distributions when the data sets were analysed separately, are also important when the entire suite of physico-chemical variables is used to constrain the analysis. Maximum lake depth, maximum altitude, TOC and Ca are all strongly correlated with the first RDA axis, with NO3 and NCR being correlated with the second RDA axis.

Forward selection (999 Monte Carlo permutation tests, Bonferroni-type corrected p-values) of the environmental variables suggests that just three of the fifteen physico-chemical variables explain statistically significant ($\alpha = 0.01$) amounts of extra variance (Table 28).

Table 28 Results of the forward selection of the 24 physico-chemical variables. Only those variables that explain statistically significant additional amounts of variance in the Cladocera data set are shown. Statistical significance assessed at α=0.05 using 999 Monte Carlo permutations tests. P-value is the significance of the F-value assessed by permutation tests. P-required is a Bonferroni corrected P-value. λ = Eigenvalue for the environmental variable and is a measure of the (extra) variance explained. % is the extra variance explained expressed as a percentage.

Variable	F	P-value	P-required	λ	%
Max Lake Depth (MLDepth)	4.420	0.001	0.05	0.052	5.2
Calcium (Ca)	3.563	0.0010	0.025	0.04	4.0
Total Aluminium (Altot)	2.144	0.0160	0.017	0.024	2.4

These three variables explain a total of 11.6% of the variance in the cladoceran data with the majority of the variance (9.8%) explained on the first and second RDA axes, the third constrained axis is trivial, explaining little of the variance in the species data. Forward selection may not selected the 'best' minimal adequate model, indeed there may not be a minimal adequate model. Many models may exist that explain similar amounts of variance that are statistically significant, yet the order in which the selection procedure chooses the important components will result in only one of these models being evaluated. To investigate which of the 24 physico-chemical variables were important in explaining the variance in the cladoceran data, a series of ordinations were performed on sets of physico-chemical variables and on individual variables.

Each of the 24 physico-chemical variables was used separately as the only environmental variable used in an RDA of the cladoceran data. The remaining 23 environmental variables were deleted from each ordination. The first axis of this ordination was then constrained by the single environmental variable, with subsequent axes being unconstrained principal components. The first axis of these ordinations was tested using 999 Monte Carlo permutation tests to determine whether the variable being tested explained statistically significant amounts of the variance in the species data. The ratio between the first, constrained axis, and the second, unconstrained axis was calculated. A high ratio indicates that a variable is important in explaining the species data. The results of theses analyses are shown in Table 29. Of the 24 physico-chemical variables, only four of the variables were statistically significant and the 99.9% level; Ca, TOC, MaxAlt and MLDepth. A further three variables were significant at the 99.5% level; EqAlk, pH and NCR.

These variables were included as the only environmental variables in and RDA. The 7 variables explain a total of 19.6% of the cladoceran species data, approximately half of the total variance explained by the 24 physico-chemical variables. Some of these variables are correlated with each other, e.g. pH and EqAlk, as indicated by high (VIF >10), so this perhaps is not the best model we can create as some degree of redundancy still remains.

Table 29: Results of individual ordinations using a single variable from the 24 physico-chemical variables. Shaded p-values indicate those variables that explain statistically significant amounts of the variance in the cladoceran data (α=0.001), except those marked with '', which are significant at α=0.005.*

Variable	λ_1	λ_1/λ_2	% variance	p-value
Alkalinity [Alk1]	0.017	0.099	1.7	0.116
Labile aluminium [Allab]	0.022	0.129	2.2	0.036
Monomeric Aluminium [Almon]	0.018	0.105	1.8	0.111
Total aluminium [Altot]	0.025	0.146	2.5	0.016
Calcium [Ca]	0.047	0.290	4.7	0.001
Chloride [Cl]	0.026	0.152	2.6	0.011
Conductivity [Cond]	0.023	0.139	2.3	0.032
Equivalent Alkalinity[EqAlk]	0.031	0.183	3.1	0.004*
Potassium [K]	0.026	0.155	2.6	0.016
Magnesium [Mg]	0.020	0.117	2.0	0.037
Sodium [Na]	0.022	0.129	2.2	0.037
Nitrate [NO3]	0.017	0.1	7.	0.146
pH	0.036	0.214	3.6	0.002*
Sulphate [SO4]	0.018	0.107	1.8	0.112
Total Organic Carbon [TOC]	0.045	0.0285	4.5	0.001
% Afforestation [Afforest]	0.015	0.088	1.5	0.259
Area to Depth Ratio [Area:Dep]	0.023	0.136	2.3	0.023
Catchment to Lake Area ration [CA:LA]	0.01	0.058	1.0	0.646
Catchment Area [CArea] (ha)	0.013	0.077	1.3	0.328
Lake Altitude [LAlt] (m)	0.029	0.173	2.9	0.010
Lake Area [LArea] (ha)	0.017	0.101	1.7	0.119
Maximum Altitude in Catchment [MaxAlt] (m)	0.040	0.248	4.0	0.001
Max Lake Depth [MLDepth] (m)	0.052	0.315	5.2	0.001
Net Catchment Relief [NCR] (m)	0.034	0.214	3.4	0.002*

An alternative approach to selecting a minimally adequate (i.e., the simplest model that still adequately explains the variance in the data set) set of environmental variables might be to remove sequentially those variables that have high VIFs when all variables are used in the RDA. After removing the variable with the largest VIF the RDA is run again the variable with the next largest VIF is removed. The process is continued until none of the variables has a VIF greater than 20. Forward selection can be performed on the remaining variables to select a model that may have greater explanatory power than the original model as chosen using forward selection. Cl, pH and CArea were removed sequentially, resulting in no other variable having a VIF greater than 20. This model explained 37.1% of the total variance in the dataset out of a possible 40.5%.

Forward selection was then used on the remaining 21 physico-chemical variables to select a minimal adequate model for explaining the variance in the cladoceran data. Three variables were found to explain statistically significant (999 Monte Carlo permutations, Bonferroni corrected p-values) amounts of extra variance; MLDepth ($\lambda_{MLDepth}$ = 0.052), Ca (λ_{Ca} = 0.040) and Altot (λ_{Altot} = 0.024). These three variables were included in an RDA, which confirmed that a total of 11.6% of the variance in the cladoceran data was explained.

It may be possible to find a model that sits between the model described above, and the 7 variables selected from Table 29. The nature of acid lakes dictates that Ca is often negatively correlated with aluminium concentrations in the lake water. Those sites which have either acidified or are naturally acidic tend to have low acid-buffering capacities and, therefore, low lake water calcium values. At low pH aluminium becomes more biologically available and higher concentrations are found. Therefore, most of the variance in the model explained by Altot is probably explained by Ca. As noted above, forward selection is often dependent upon the order in which the variables are selected. A number of permutations of variables chosen during forward selection were tried to see the effect they had on the model. MLDepth and Ca were included in the model first, as they explained twice as much additional variance as any other variable in the training set. Altot was not included in the model and in its place another variable that explained a similar amount of extra variance was included. This was tried repeatedly, each time a different variable was chosen as the first variable in the model and forward selection was continued until no further variables added statistically significant amounts of extra explained variance to the model.

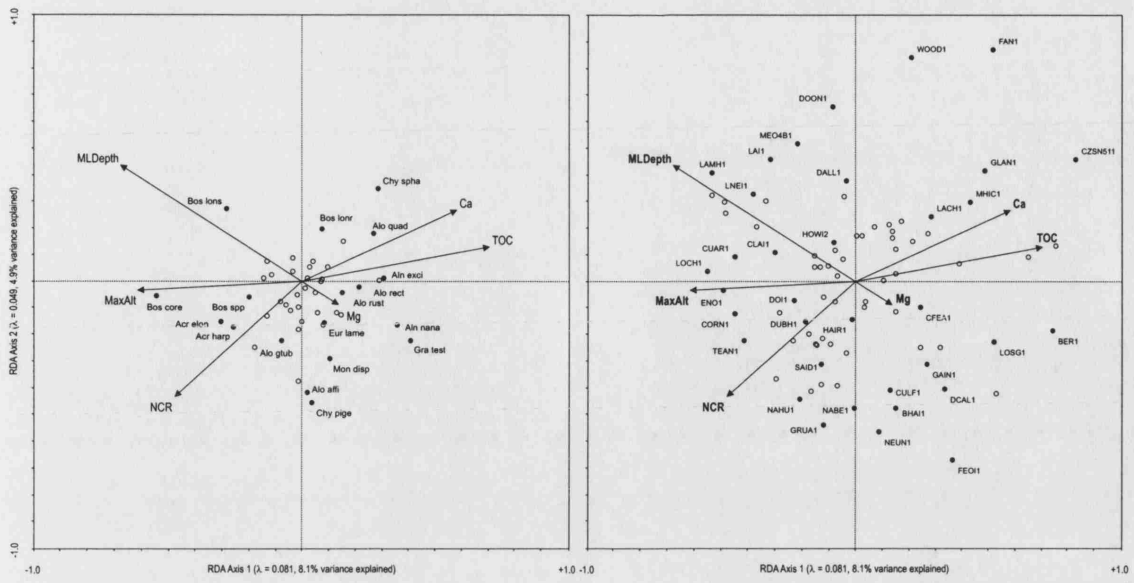


Figure 43: Redundancy analysis biplots showing the final model of physico-chemical variables.

The variable with the next highest extra variance explained after Altot was K. K was shown not to explain a significant amount of variance on its own (See Table 29) and the resulting model performed less well than the others tried previously. After K, NCR explained the next largest amount of extra variance explained. NCR also explained a statistically significant amount of variance in the cladoceran data when used as the only environmental variable in RDA (See Table 29). NCR was included as the third variable in the model and forward selection was continued. The results of this model are shown in Table 30 and Table 31. This model explained 19.3% of the variance in the cladoceran data, 0.6% less than the 7-variable model described above. Furthermore, the variance inflation factors for the 6 variables in the model are all very low indicating that the variables are not particularly correlated, and therefore are explaining independent amounts of the variance in the cladoceran data. The first axis of the ordination and the sum of all canonical eigenvalues (i.e. trace of the matrix of eigenvectors) are significant at $p = 0.001$ as assessed by 999 Monte Carlo permutation tests. A biplot of this model is shown in Figure 43.

Table 30: Results of the forward selection of environmental variables to choose a minimally adequate model.

Variable	F	P-value	P-required	λ	%
Max Lake Depth (MLDepth)	4.420	0.001	0.05	0.052	5.2
Calcium (Ca)	3.563	0.001	0.025	0.04	4.0
Net catchment relief	1.994	0.013	0.017	0.022	2.2
Maximum altitude	2.370	0.003	0.0125	0.026	2.6
Total organic carbon	2.365	0.005	0.01	0.026	2.6
Magnesium	2.530	0.003	0.008	0.027	2.7

Table 31: Summary statistics of the final RDA model using 6 physico-chemical parameters as explanatory variables.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.081	0.049	0.029	0.020	1.000
Species – Environment correlation	0.718	0.729	0.686	0.523	
Cumulative % variance of species data	8.1	13.1	15.8	17.8	
Cumulative % variance of species environment correlation	42.2	67.6	82.0	92.3	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.193

The lakes that form the reduced training set are taken from a large geographic range spanning Wales and Scotland. Some of the variance in the cladoceran data may, therefore, be spatially structured. It is important to look at this spatial structure, to investigate whether the variation in the Cladocera that is attributed to variations in physico-chemical properties of the lakes in which they were living can be explained by geographic distances alone.

In order to assess the spatial structure in the data, a 3rd order polynomial was created from the centred UK national grid references of the site locations. The 3rd order polynomial describes a complex spatial trend surface and contains nine variables. These nine variables

were; X, Y, XY, X^2Y, XY^2, X^2, Y^2, X^3 and Y^3. To reduce the effects of scale these variables were standardised before being analysed with RDA and partial RDA.

RDA of the cladoceran species data, with the spatial variables used and explanatory variables, showed that 17.7% of the variance in the species data could be explained by spatial differences between sites (Table 32). Approximately half of this variance is explained by the first axis of the RDA, with small amounts of variance explained on subsequent axes. The results indicate that there is some clear underlying spatial structure to the cladoceran data. A forward selection procedure showed that of the nine spatial variables only three, Y^3, Y and XY, explained statistically significant amounts of the variance in the cladoceran data, accounting for 10.4% of the variance in the species data.

To explore the relative strengths of the spatial, physical and hydrochemical parameters in explaining variance in the cladoceran abundance data the 33 spatial, physical and hydrochemical variables were used as explanatory variables in RDA. 51.8% of the variance in the cladoceran data was explained by the 33 spatio-physico-chemical variables Table 33.

Table 32: Results of the RDA of the cladoceran data with the nine spatial parameters used as explanatory variables.

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.061	0.034	0.028	0.015	1.000
Species – Environment correlation	0.688	0.541	0.626	0.551	
Cumulative % variance of species data	8.1	9.4	12.2	13.7	
Cumulative % variance of species environment correlation	34.4	53.4	69.1	72.5	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.177

Forward selection showed that Y^3 and XY explained an extra 4.2% and 2.9% of the variance in the cladoceran data, and were the second and fourth most important environmental variables of the 33 analysed. As a comparison, MLDepth and Ca were the only other variables to explain statistically significant amounts of extra variance, and explained 5.2% and 3.7% of the variance in the cladoceran data respectively. Together,

these four variables explained 16% of the total variance in the cladoceran data. These results clearly show that some of the cladoceran data is spatially structured.

Partial ordination techniques analyse the residual variance of an ordination of the species data and a number of covariables used as explanatory parameters, and the environmental variables of interest. The end result is an ordination of the variables of interest after the variance explained by the nuisance or unimportant variables has been taken into account.

The spatial structure in the cladoceran data is unimportant; rather it is the variation in the cladoceran communities with respect to the physico-chemical parameters that it is important to explain. To evaluate the effects of the physico-chemical parameters after the spatial structure was removed, a partial RDA (pRDA) using the 3 significant spatial variables as covariables and the 24 physico-chemical variables as explanatory variables was performed. The remaining 6 spatial variables were not included in the analysis. The results of the pRDA are shown in Table 34.

Table 33: Results of the RDA of the cladoceran data with the 33 spatio-physico-chemical parameters used as explanatory variables.

Axes	**1**	**2**	**3**	**4**	**Total Variance**
Eigenvalue (λ)	0.127	0.071	0.051	0.048	1.000
Species – Environment correlation	0.885	0.759	0.789	0.802	
Cumulative % variance of species data	12.7	19.8	24.9	29.7	
Cumulative % variance of species environment correlation	24.6	38.2	48.1	57.3	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.518

Table 23 shows that the 24 physico-chemical variables explained 40.5% of the variance in the cladoceran data. After partialling out the spatial structure in the species data, the physico-chemical variables explained 34.9% of the variance (Table 34). The inter-set correlations of the pRDA show that both Ca and TOC are negatively correlated with the first axis of ordination, MLDepth with the second, with the third axis being an acidity gradient correlated with pH, EqAlk and Altot.

In previous analyses (see above) six physico-chemical variables were used to explain statistically significant amounts of the variance in the cladoceran data. A pRDA (3 spatial covariables, 6 physico-chemical explanatory variables) was performed to remove the spatial structure from the cladoceran data before applying this explanatory model. The six physico-chemical variables explained 15.3% of the variance in the cladoceran data, two-thirds of which (10.7%) was explained by the first two axes of variation.

Monte Carlo permutations (999 permutations) were used to assess the significance of the first and all canonical eigenvalues. Both the first canonical eigenvalue (F-ratio = 5.099, p = 0.001) and all canonical eigenvalues (F-ratio = 2.509, p = 0.001) were shown to be statistically significant.

Table 34: Results of the pRDA, 3 spatial covariables and 24 physico-chemical explanatory variables, of the cladoceran data.

Axes	**1**	**2**	**3**	**4**	**Total Variance**
Eigenvalue (λ)	0.090	0.052	0.044	0.029	1.000
Species – Environment correlation	0.830	0.695	0.780	0.697	
Cumulative % variance of species data	10.0	15.8	20.7	24.0	
Cumulative % variance of species environment correlation	25.7	40.6	53.2	61.6	
∑ all unconstrained λ					0.896
∑ all canonical λ					0.349
∑ λ for covariables					0.104

The analyses presented above illustrate the influence of biogeographical factors on the distribution of Cladocera in UK upland lake systems. Furthermore, they illustrate the relative importance of physical, hydrochemical and spatial factors in determining the species distributions.

Variance partitioning, proposed by Borcard *et al.* (1992), is an ordination technique where the analyst can determine the amounts of unique variance explained by an environmental

variable or groups of variables. Interactions between environmental variables or sets of variables can also be determined. In this way, one can partition the variance in species distributions into the variance explained uniquely by the environmental variables or set of variables, and the variance that is conditional on other variables or sets of variables.

Variance partitioning of the cladoceran species data was performed using three sets of environmental variables; physical, hydrochemical and spatial. The 33 environmental variables in the data set were divided into these classes as shown in Table 35. A series of RDA and pRDA were performed to partition the variance between the three sets of environmental variables. The results are shown in Table 36 and Figure 44. Table 36 shows the calculations used in the variance partitioning process.

Table 35: Listing of the three sets of explanatory variables used in the variance partitioning of the cladoceran data.

Physical	Hydrochemical		Spatial
% Afforestation [Afforest]	Alkalinity [Alk1]	Magnesium [Mg]	X
Area to Depth Ratio [Area:Dep]	Labile aluminium [Allab]	Sodium [Na]	Y
Catchment to Lake Area ration [CA:LA]	Monomeric Aluminium [Almon]	Nitrate [NO3]	XY
Catchment Area [CArea] (ha)	Total aluminium [Altot]	pH	X^2
Lake Altitude [LAlt] (m)	Calcium [Ca]	Sulphate [SO4]	Y^2
Lake Area [LArea] (ha)	Chloride [Cl]	Total Organic Carbon [TOC]	X^2Y
Maximum Altitude in Catchment [MaxAlt] (m)	Conductivity [Cond]		XY^2
Max Lake Depth [MLDepth] (m)	Equivalent Alkalinity[EqAlk]		X^3
Net Catchment Relief [NCR] (m)	Potassium [K]		Y^3

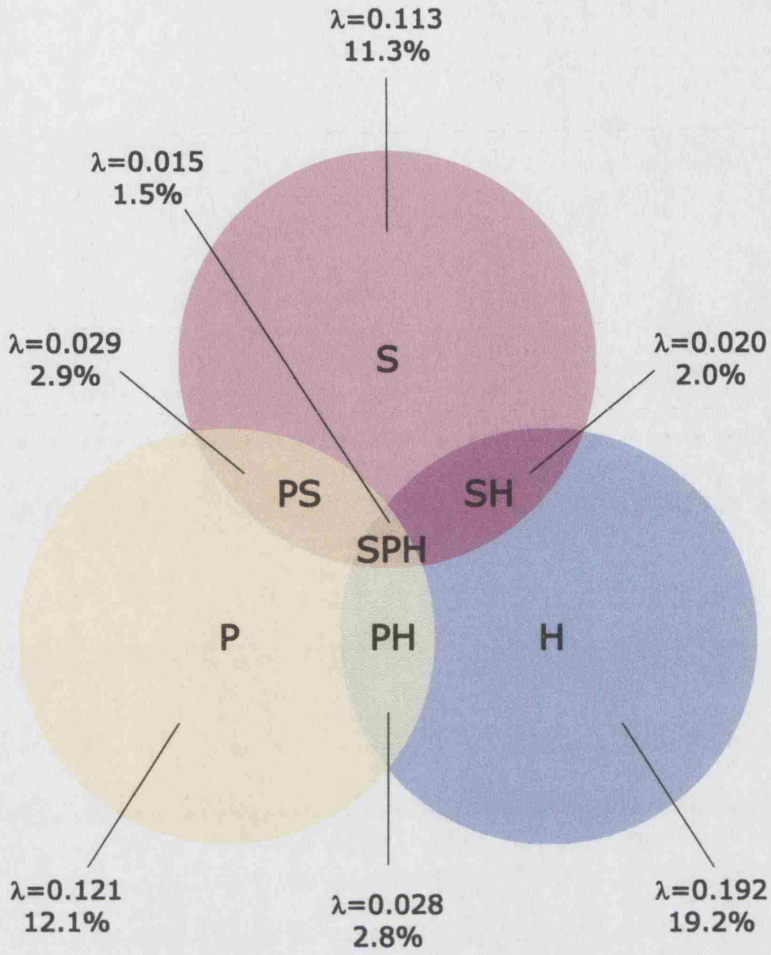


Figure 44: Venn diagram showing the results of the variance partitioning exercise for the 83-lake cladoceran training set. S = spatial variables, P = physical variables, H = hydrochemical variables. The variance of each section is presented as the eigenvalue (λ) and the percentage of the total variance (= 1.000) explained. Variance that is unique to a set of variables is described by the section of each circle not overlapped by another circle. Variance conditional on another set of variables is indicated by the overlap of two or more circles.

In Figure 44 the conditional variance is shown as the overlap of the circles between two sets of variables (areas PS PH, SH) or all three sets of variables (area SPH). The main feature of Figure 44 is that very little of the variance in the cladoceran data explained by a single set of variables is conditional on the other two sets of variables. This confirms the hypothesis from previous analyses presented earlier that the results indicated that the variables were explaining unique amounts of variance in the species data.

Table 36: Calculations used in the variance partitioning of the Cladocera data

Area	Environmental Variables	Covariables	λ	%
S	S	P+H	0.113	11.3
P	P	S+H	0.121	12.1
H	H	S+P	0.192	19.2
SP	S	H	0.029	2.9
SH	H	P	0.020	2.0
PH	P	S	0.028	2.8
S+PS+SPH+SH (S_{total})	S	-	0.177	17.7
SPH	= S_{total} – SP – SH – S		0.015	1.5
Variance Explained			0.518	51.8
Unknown Variance			0.482	48.2
Total Variance			**1.000**	**100.0**

The hydrochemical variables explain the greatest amount of the variance in the cladoceran species data, accounting for 19.2% of the variance alone. A further 6.3% of the variance is explained by unique interactions between the hydrochemical variables and the physical (2.8%) and the spatial variables (2.0%), and the unique interaction of all three sets of variables (1.5%). The physical variables explain an additional 12.1% of the variance, with unique interactions between the other sets of variables accounting for a further 7.2% of the variance. The spatial variables account for the smallest amount of variance of the three sets of variables, accounting for 11.3% alone, with a further 6.4% of the variance explained by unique interactions between the spatial variables and the other sets of environmental variables.

4.3 Cladoceran species response curves

The previous section of the thesis assessed the importance of the measured physico-chemical data in explaining the main patterns in the entire cladoceran training set. The approach used above produces a low dimensional representation of the individual species responses to each of the environmental variables and as such approximates the real taxon

responses to the environmental data. This section of the thesis will focus on individual species responses to selected environmental variables using generalised linear modelling techniques.

Modelling species-environment relationships can be thought of as modelling the realised niche; the environmental space where a species can coexist with other species. These models can be used to explain species distributions, to perform environmental reconstructions from palaeoecological data, and to predict future changes in species distributions, for example under climate change scenarios.

Much cladoceran autecological work has been performed through field observations or laboratory manipulation studies. Where large scale spatial studies have been performed, subsequent analysis of the species and physico-chemical data has been restricted to multivariate methods. Few, if any, univariate statistical analyses have been performed on large species data sets. This is in contrast to other aquatic groups including the widely studied diatoms (Birks *et al.* 1990b; Stevenson *et al.* 1991; Dixit *et al.* 1993; Cameron *et al.* 1999) or aquatic macrophytes. The direct result of this paucity of autecological data for the Cladocera is that the factors controlling the distribution of cladoceran species are not well known and subject to much controversy.

Multivariate statistics provide a holistic view of the species-environment relationship. Species and sample points are analysed as a whole and the main patterns in the species data are extracted as latent variables or gradients. Using direct gradient analysis methods (e.g. RDA and CCA) these latent variables can be constrained to be a linear combination of the measured environmental data to relate the main patterns in the distributions of the species to the main patterns in the environmental data. These multivariate methods allow the user to identify the broad patterns in the data and to approximate the response of individual taxa to the measured environment. If the aim of the analysis is to relate individual species' abundances to the measured environmental data a more traditional, univariate approach is required.

In a univariate analysis each species is dealt with in turn by a separate regression. Traditionally a single environmental variable is used as the predictor variables and the abundance of a particular taxon is regressed upon this predictor. Multiple regression

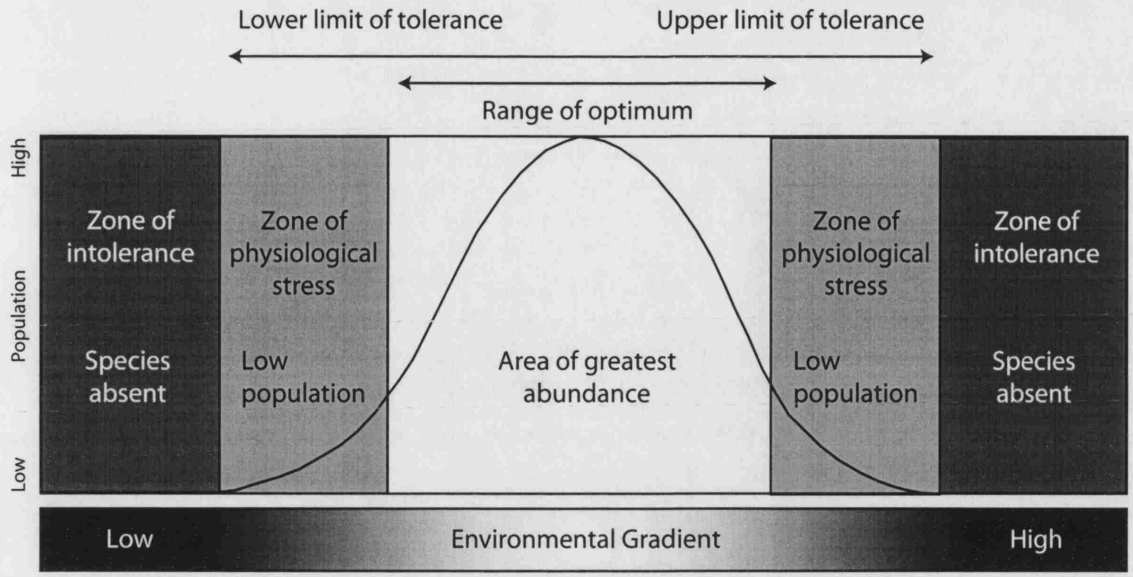


Figure 45: The Gaussian or normal curve of a species' response to a single environmental factor and zones of tolerance. Redrawn from §1.6 in Kent and Coker (1992).

procedures can be used to relate species abundances to more than just a single predictor, though this type of approach can quickly become complicated to interpret ecologically and difficult to visualise, and a multivariate approach might be more appropriate.

If one were to take a single environmental factor and plot the abundances of a single taxon at sampling locations along a gradient of values for that environmental factor the resulting plot would likely conform to a unimodal or normal curve if the sampled gradient is sufficiently long. This concept is illustrated in Figure 45. The area along the environmental gradient where the species is most abundant is likely to be where optimal environmental conditions for this species are found. Either side of this optimal area the species is likely to be present but at low abundances due to physiological stress caused by the non-optimal conditions and competition with other species more suited to those conditions. Outside this tolerance range the environmental conditions are unsuitable for the species to exist and the species is observed to be absent from these locations. Work by Whittaker (e.g. 1956; 1967) has shown that species abundances generally

Logit regression is a specialised form of the Generalised Linear Model (GLM, see Section 2.4.5.1 above) and can be modelled using a binomial error distribution and a logit link function. The Gaussian logit curves presented in this section were all modelled using R version 1.5.1 (Ihaka and Gentleman 1996), using maximum likelihood estimation by an

iterative weighted least squares procedure (Venables and Ripley 1999). The species abundance data (percentages) were transformed to presence/absence data prior to analysis. The graphical output from the analysis presented in the subsequent pages follows a standard form.

Each species presence/absence observation for a given value of x, the environmental variables, is presented as a point along the top of a plot if the taxon is present in the sample and along the bottom if absent. The Gaussian logit regression (GLR) response curve is plotted as a solid line through the data points and the $\pm$ 1 standard error is shown by the dashed lines. The x-axis represents the environmental variable used in each analysis and the units of measure are stated.

Appendix D contains the full range of fitted response curves for each of the 23 physicochemical variables for all taxa in the training set. Significant or interesting response are highlighted and expanded upon below.

Detrended (constrained) correspondence analysis of the cladoceran data has shown that the species data exhibit short gradient lengths with respect to both the latent, hypothetical gradients and the measured environmental data. It is interesting to note that the results of the GLR analyses show that the species show a mixture of linear, monotonically rising or falling and unimodal response to single environmental variables. Clearly, for certain taxa the data set covers the whole range of physiological tolerance, whilst for other species only part of this range is captured by the data. Table 29 above shows the results of many ordinations using a single environmental variable as the constraint in the RDA. Maximum lake depth, when used as a single explanatory, explains 5.2% of the variance in the cladoceran data. Figure 46 shows the GLR models for four chydorid species (*Graptoleberis testudinaria*, *Alona intermedia*, *Rhynchotalona falcata* and *Chydorus sphaericus*) and maximum lake depth. *G. testudinaria* is generally absent from the deeper lakes in the training set, being found in only 4 lakes with a maximum depth greater than 25 metres. This is reflected in the shape of the response curve for this taxon, which indicates a high probability of occurrence only in lakes with a maximum depth less than 10 metres. The occurrence of *G. testudinaria* is strongly related to the presence of and abundance and type of aquatic macrophytes in a lake (Quade 1969). This taxon is a macrophyte associated species often found on *Potamogeton*, *Ceratophyllum*, *Myriophyllum*, *Elodea* and *Phragmites*. Fryer (1968, page

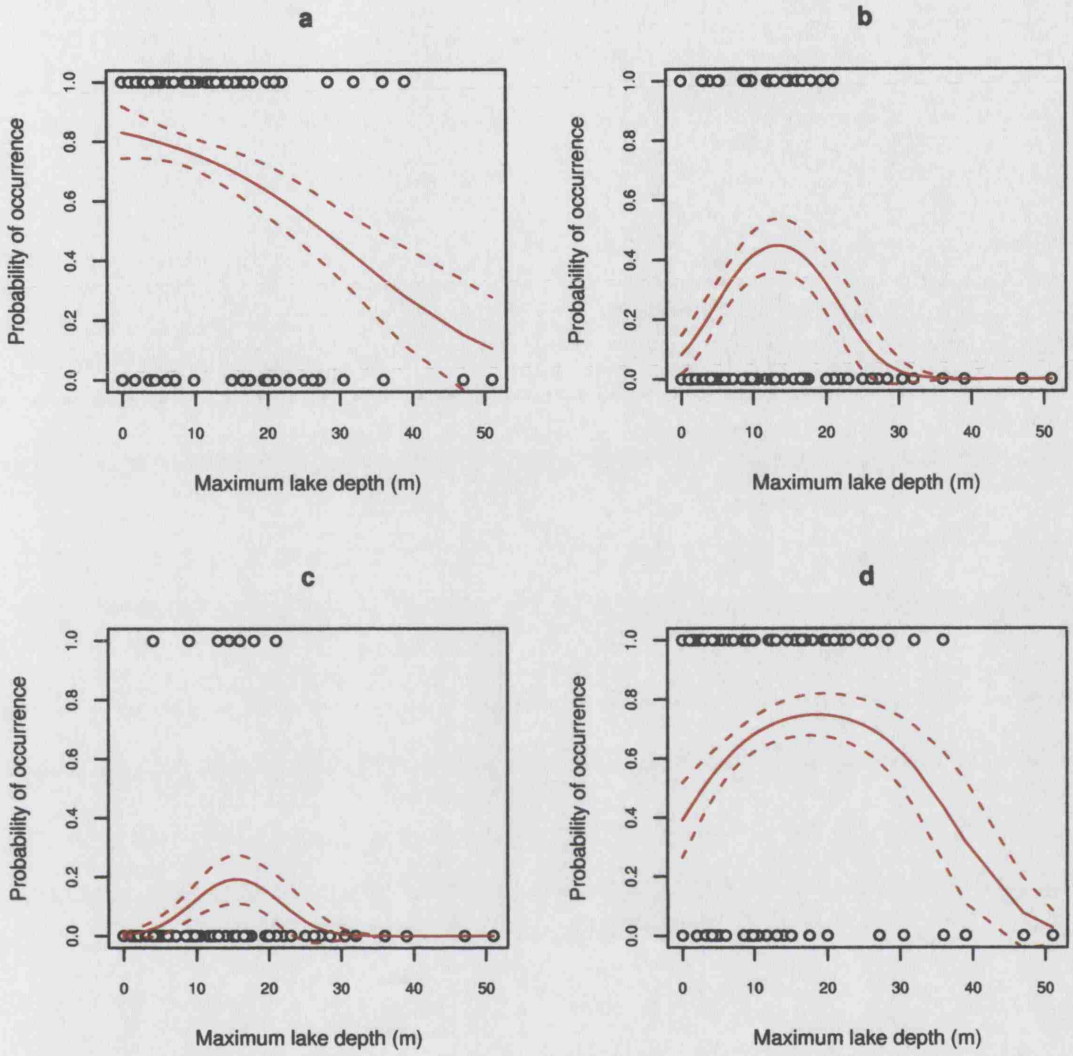


Figure 46: Gaussian logit regression models for four species, G. testudinaria *(a),* A. intermedia *(b),* R. falcata *(c) and* C. sphaericus *(d), showing the relationships between species presence/absence and maximum lake depth (in metres). The points represent presence (1) or absence (0) of the taxon. The dashed lines are the ± 1 standard errors of the fitted values (solid line).*

222) even describes *G. testudinaria* as "*..., a chydorid with the habits of a gastropod slug*", in reference to his observations of this species gliding over the surfaces of plants collecting food with its second trunk limbs. Whilst maximum lake depth is clearly only a poor surrogate for the type, amount or extent of aquatic macrophyte habitat available in lakes, it is clear that the availability of suitable habitats for *G. testudinaria* is inversely correlated with lake depth.

A. intermedia shows a quite restricted range about an optimum of 10 metres water depth. Indeed, this taxon is not found in lakes deeper than 20 metres. *A. intermedia* is known in Britain from predominately Scottish locations, often in weakly acidic sites in the boreo-

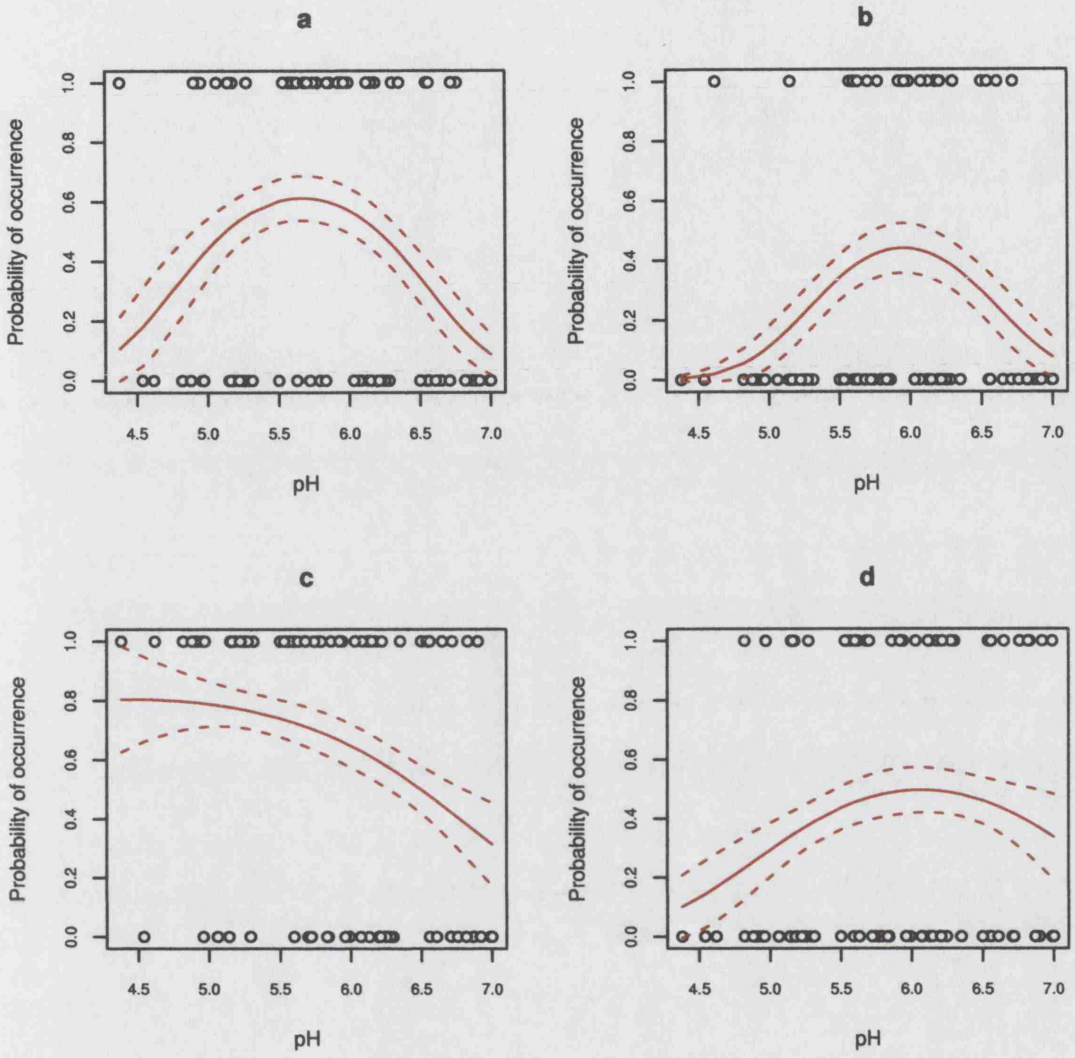


Figure 47: Gaussian logit regression models for four species, A. guttata *var.* tuberculata *(a),* A. intermedia *(b),* M. dispar *(c) and* S. crystallina *(d), showing the relationships between species presence/absence and pH. The points represent presence (1) or absence (0) of the taxon. The dashed lines are the ± 1 standard errors of the fitted values (solid line).*

alpine zone (Fryer 1993). The species is associated with the benthos, moving over muddy detrital substrates close to the margins of the water body. It is often found in moorland sites and as such, these sites are not likely to be particularly deep or have a quickly shelving bottom.

The response curve for *R. falcata* is also similar to that of *A. intermedia*, with this species only being found in those training set lakes with a maximum depth of 20 metres or less. *R. falcata* is known to have specific habitat preferences, and is most often (though not uniquely) found inhabiting the sandy substrates of upland, oligotrophic, acidic moorland

systems and this is again reflected in the maximum depth response curve for this species. The response curves for other environmental variables reflect these known physiological preferences, indicating increased probabilities of occurrence in lakes with low ionic strength (conductivity, magnesium and calcium) and at intermediate altitudes.

Chydorus sphaericus is a strange chydorid in that, unlike most other chydorids it is a strong swimmer, and is often found in both the open littoral and the open pelagic zones of lakes, though its presence in the pelagic zone is often related to the presence of filamentous algae to which *C. sphaericus* attaches itself (Fryer 1968). The species is found most abundantly in lakes approximately 20 metres deep and in the training set is represented in lakes that are up to 37 metres in depth. *C. sphaericus* is equally well represented in the shallower water bodies, as is indicated by the wide tolerance range for this taxon, as illustrated in Figure 46.

Other interesting results from the GLR analyses include the species-pH relationships. *Bosmina longirostris* is largely absent from the training set, being present in only a 5 samples (Max. abundance 15.38%). Two other members of the genus *Bosmina*, *B. longispina* and *B. coregoni*, conversely are present in the majority of samples. Interestingly, the only samples that *B. coregoni* is absent from are found at the upper end of the pH range for the training set, above pH 6.0. The pH-response curve for *B. coregoni* might indicate that the training set covers some sites in the range of physiological stress for this particular taxon but the extent of the training set is insufficient to capture the full response curve for this taxon. None of the species found in the training set shows a particular restriction to the very acidic lakes. *Alona guttata* var. *tuberculata* is most abundant in acidic waters around a pH of 5.5-5.75, but is present across the acidity gradient and exhibits a wide tolerance range with respect to pH.

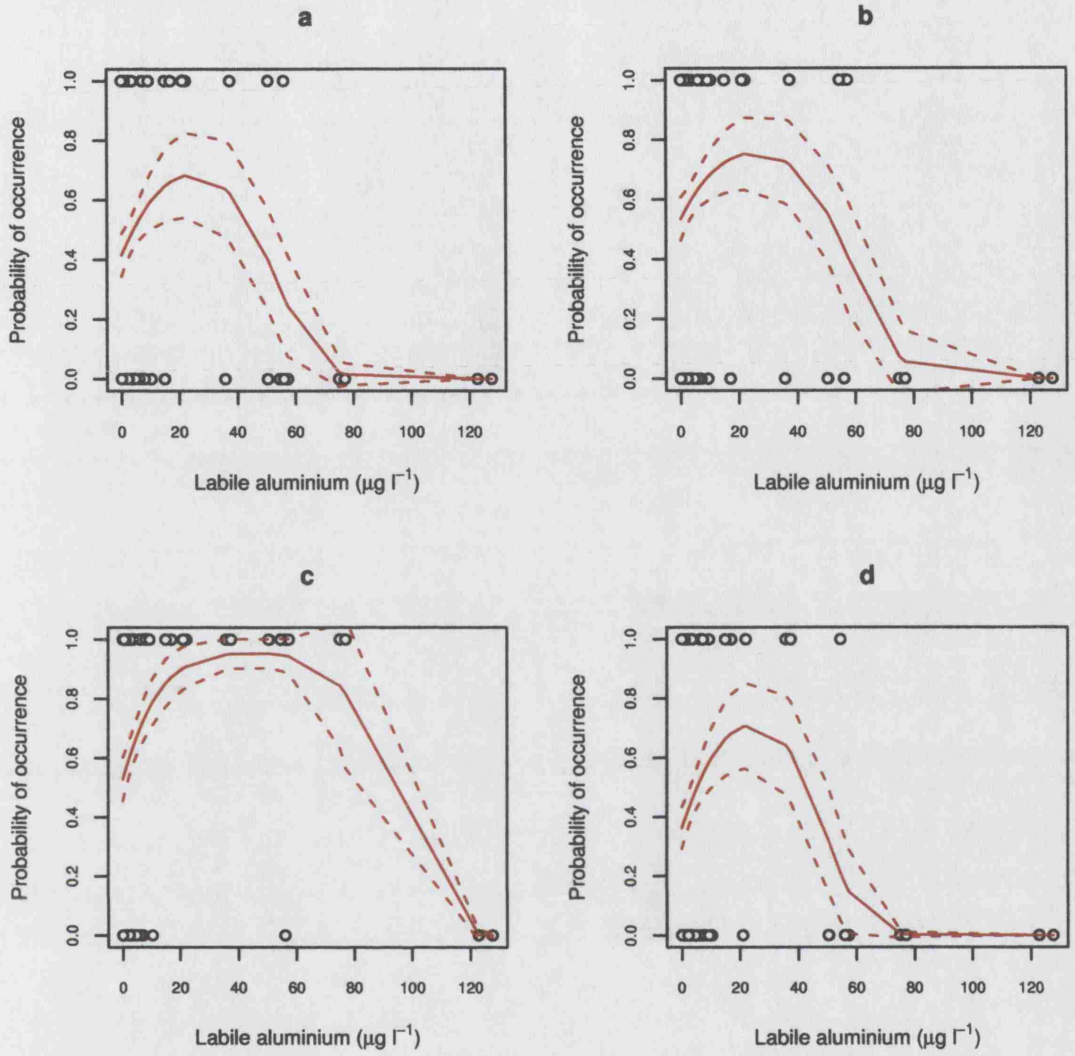


Figure 48: Gaussian logit regression models for four species, A. guttata *var.* tuberculata *(a),* A. quadrangularis *(b), M. dispar (c) and S. crystallina (d), showing the relationships between species presence/absence and labile aluminium concentrations. The points represent presence (1) or absence (0) of the taxon. The dashed lines are the ± 1 standard errors of the fitted values (solid line).*

A. intermedia and *Sida crystallina* also both show unimodal responses to pH, with both taxa being most abundant around pH 6.0 in the training set. Other taxa, most notably *Alona rectangula*, *Camptocercus rectirostris* and *Monospilus dispar* show a preference for the more acidic waters; their pH response curves indicating higher probability of occurrence for waters at the lower end of the pH gradient, *ca.* pH 4.5-5.0.

The RDA results indicated an independent effect of aluminium on the distribution of the cladoceran taxa in the training set. A number of chydorid taxa show a striking unimodal response to labile aluminium concentrations in the lake water. In particular *A. guttata* var.

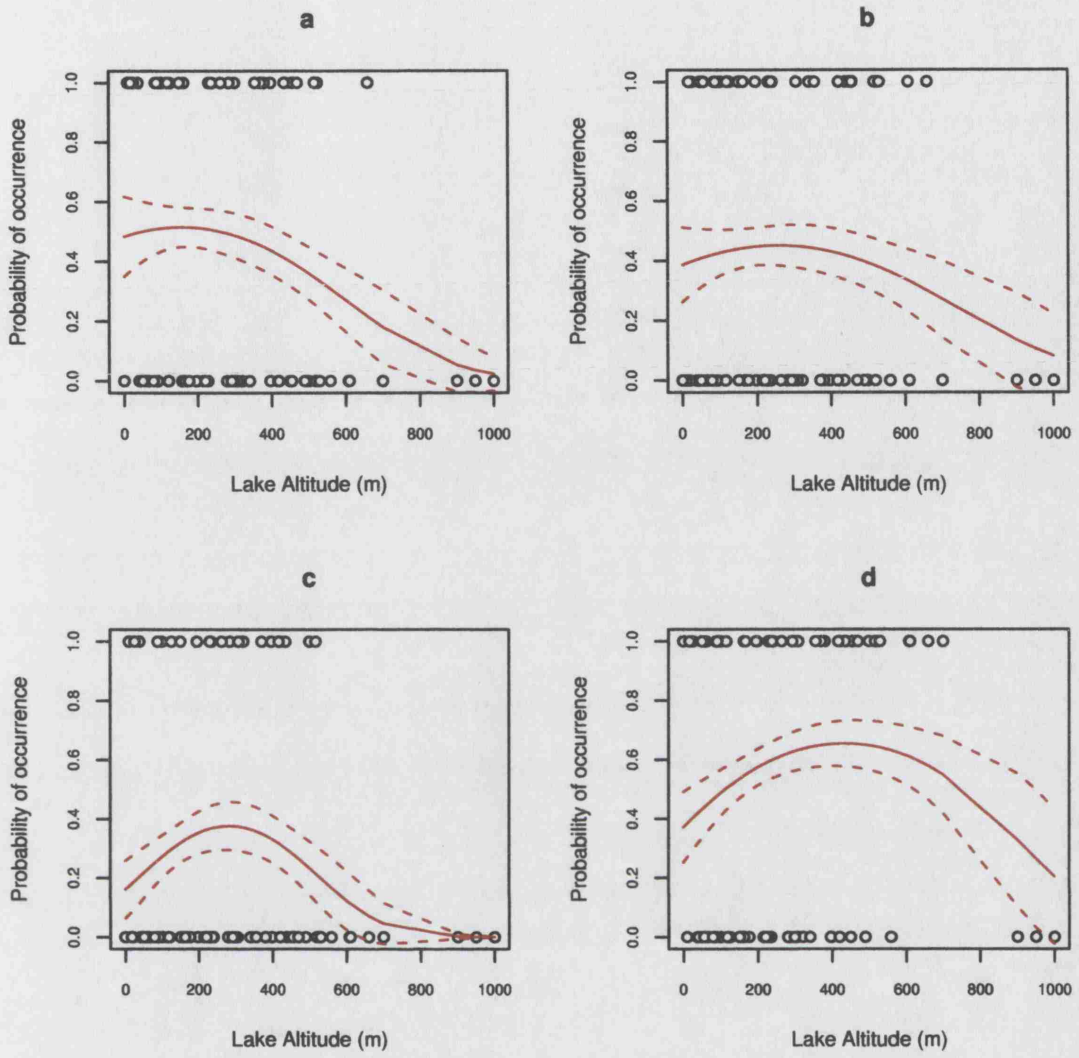


Figure 49: Gaussian logit regression models for four species, A. guttata *var.* tuberculata *(a),* A. guttata *(b),* A. intermedia *(c) and* A. quadrangularis *(d), showing the relationships between species presence/absence and lake altitude. The points represent presence (1) or absence (0) of the taxon. The dashed lines are the ± 1 standard errors of the fitted values (solid line).*

tuberculata, *Alona quadrangularis*, *M. dispar* and *S. crystallina* are restricted to waters with labile concentrations lower than 80 µg l^{-1} Figure 48.

Altitude was also shown to be an important factor in determining cladoceran species distributions across the training set. Many chydorid taxa are absent from the highest altitude sites, in particular *A. guttata* var. *tuberculata*, *Alona guttata*, *A. intermedia*, *A. quadrangularis*, *A. rectangula*, *G. testudinaria*, R. *falcata* and *S. crystallina*, and this pattern is reflected in their GLR lake altitude response curves Figure 49.

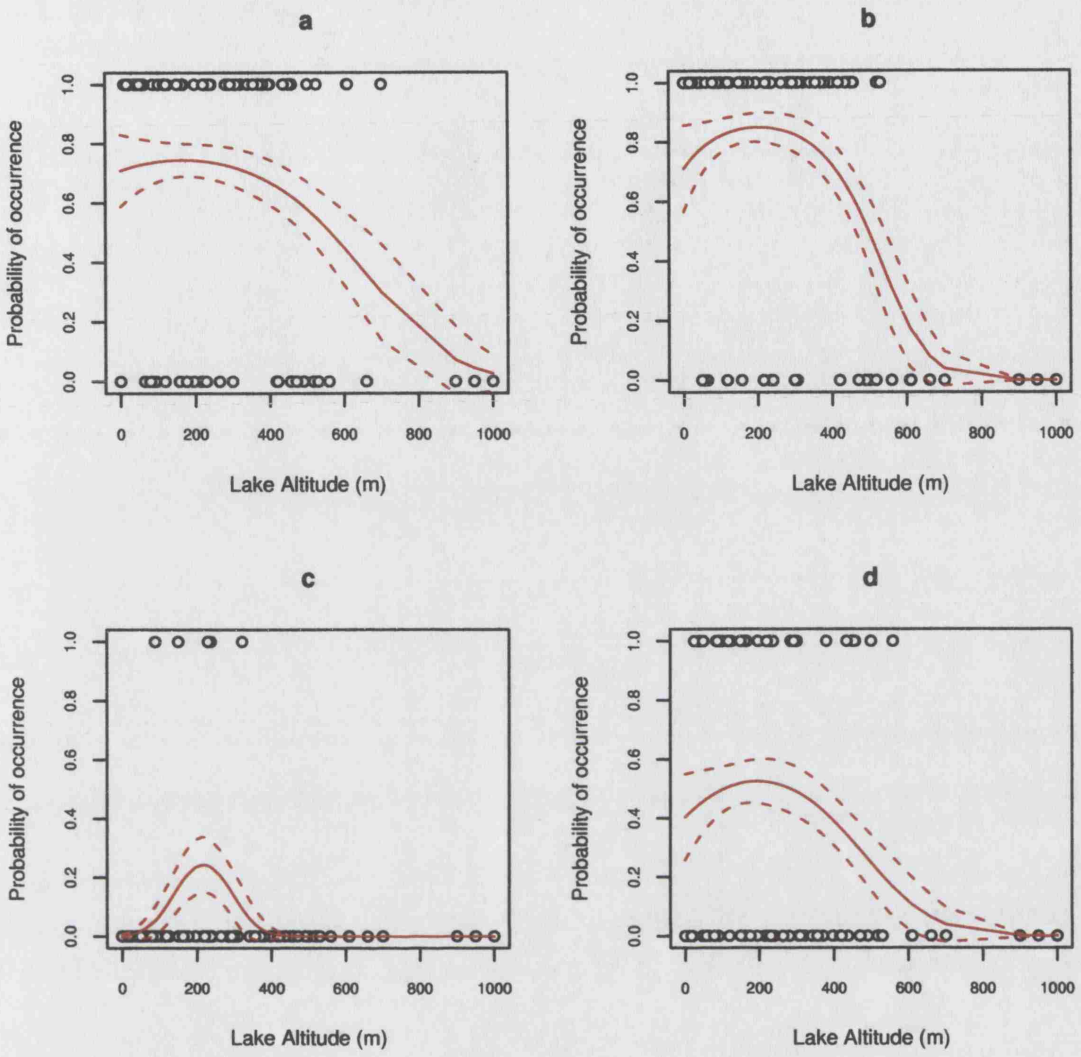


Figure 50: Gaussian logit regression models for four species, A. rectangula *(a),* G. testudinaria *(b),* R. falcata *(c) and* S. crystallina *(d), showing the relationships between species presence/absence and lake altitude. The points represent presence (1) or absence (0) of the taxon. The dashed lines are the ± 1 standard errors of the fitted values (solid line).*

Other species, such as *Acroperus harpae*, *Alonella excisa*, *Alonella nana* and *Acroperus elongata*, are known from the literature to be cold tolerant taxa (e.g. Whiteside 1970; Lotter *et al.* 1997; Hofmann 2000; Hofmann 2001) and this trait is reflected in their response to altitude across the range of lakes in the training set, being found consistently at the higher altitude sites. In the case of *A. elongata* the predicted probability of occurrence is greatest at all lakes over 400 metres, with this particular taxon present at all sites located above an altitude of 600 metres.

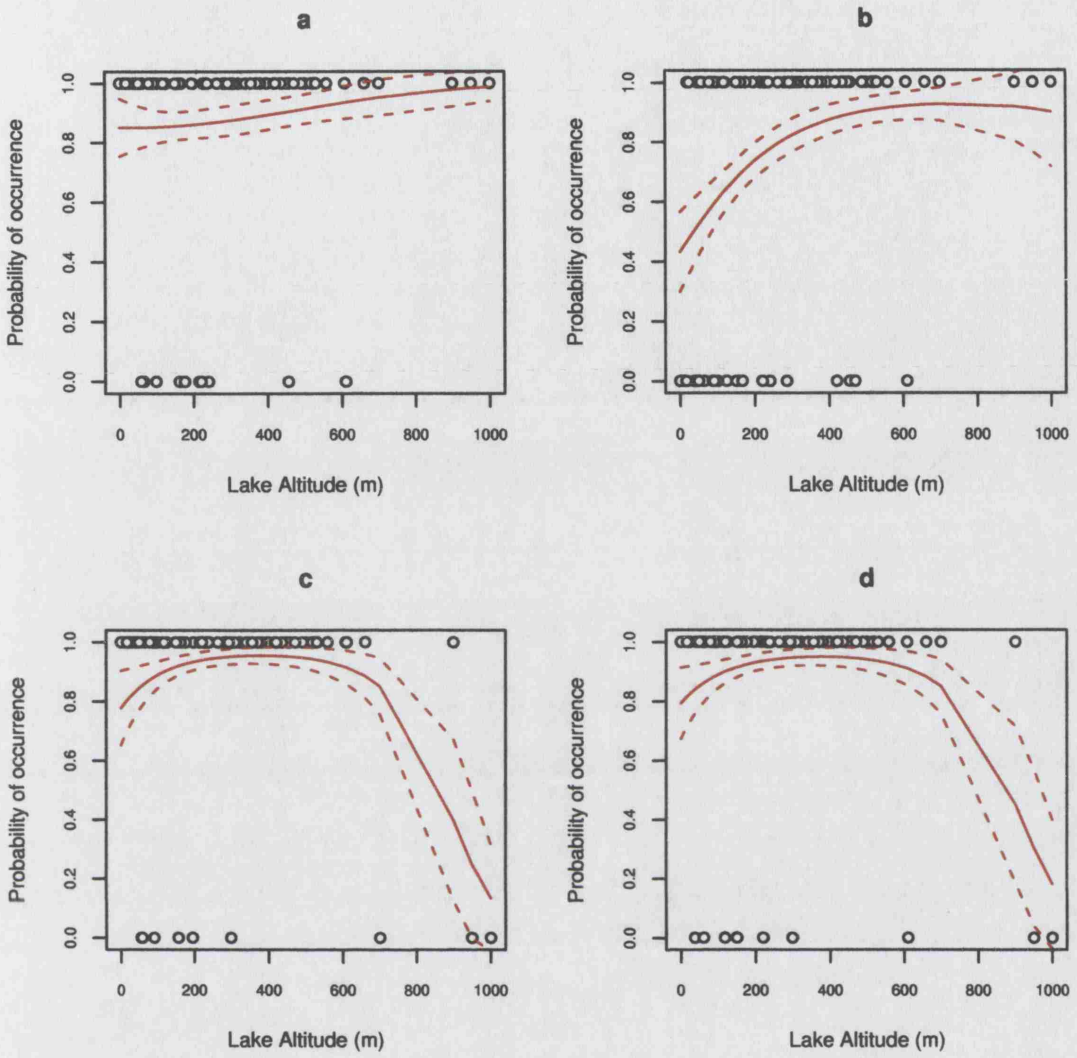


Figure 51: Gaussian logit regression models for four species, A. harpae *(a),* A. elongata *(b),* A. excisa *(c) and* A. nana *(d), showing the relationships between species presence/absence and lake altitude. The points represent presence (1) or absence (0) of the taxon. The dashed lines are the ± 1 standard errors of the fitted values (solid line).*

Appendix D contains the full range of response curves for each species in the cladoceran training set. It is clear from the range of response to the environmental data (unimodal, linear, & monotonic) that cladoceran responses to environmental gradients are incredibly complex in nature – which is one of the problems associated with a univariate approach; the shear size and complexity of ecological data sets mean that significant time and effort are required to tease out the interesting signals and patterns.

Many of the models shown in Appendix D are not statistically significant (for either the linear or the unimodal terms in the models), and this is partly due to many of the species

either being present in samples across the entire range of the environmental variable being modelled, or because a species is rare in the current data set and only occurs in a few samples. As such the results presented in this section and Appendix D should be treated with care; they should be used as a guide to the responses of the individual taxa but over interpretation of the results is cautioned against. In this sense, the response curves will be of practical use to palaeolimnologists in interpreting changes in taxon abundances through time in sediment core studies, as well as aiding the interpretation of the output from a combined diatom and cladoceran analogue matching procedure (see Chapter 5).

One area that poses significant difficulty is in the area of building simple predictive models for species – environment relationships. In this section I have used a single explanatory factor for each model, and have ignored interactions between variables leading to species response, or the possibility that two or more factors might be better at explaining cladoceran response at the individual species level. A data mining exercise in model building is beyond the scope of this chapter and this thesis, however, the use of techniques such as regression and classification trees or neural networks (e.g., Ripley 1994; Breiman 1996; Breiman 2001) may be a useful area of future work that can build upon the work presented here.

The results of the GLR presented in this section and Appendix D have highlighted some interesting species-environment relationships, particularly with lake depth, aluminium concentrations and pH, and have also suggested some inadequacies in the extent of the current training set for purposes of modelling species response curves to measured environmental data. In particular, it is difficult to determine anything about the ecological preferences of those taxa that are found in only a few samples. Many of the 48 taxa in the training set are found in only a few samples, and modelling or inferring anything about the ecological preferences and tolerances of these species from such data is impossible.

Another issue illustrated by the response curves for some of the most common taxa, in particular, *B. coregoni*, *B. longispina* and *Chydorus piger*, is that the environmental gradients in the training set are not sufficiently large to capture the full range of the species. It is almost as difficult to determine anything about the ecological preferences of these three taxa as it is for the rare taxa. The results of the GLR analyses indicate that these three taxa are ubiquitous members of the acidic, oligotrophic fresh water systems typical of upland

Britain and the captured environmental gradients do not contain sites with enough variation in nutrient status and base concentration to fully represent the ecological tolerances of these particular species.

4.4 Cladoceran community types in upland fresh waters of Scotland and Wales

The previous two sections of this chapter have provided an extensive evaluation of cladoceran species composition in oligotrophic fresh water systems from the uplands of Scotland and Wales. Section 4.2 used multivariate statistical techniques to interpret the main patterns in cladoceran species composition from 83 upland lakes and to relate these patterns to the measured physico-chemical data. Section 4.3 took a more focussed approach to analysing the species data. Species response curves to significant physico-chemical parameters were modelled using Gaussian logit regression to identify species-specific physiological requirements. These two approaches can be seen to lie at opposite ends of a scale gradient. The multivariate analyses taking a holistic, large scale approach to identifying and explaining patterns in the cladoceran data. The univariate work in Section 4.3 is more focussed, concentrating on individual species responses. In between these two extremes, we might expect to be able to identify community types within the cladoceran data, i.e. groups of species commonly found together in the same lakes. Furthermore, using discriminant analysis it may be possible to identify those physico-chemical parameters that best describe or influence these community types. This section will present the results of a cluster analysis of the cladoceran abundance data and an attempt to identify those physico-chemical parameters that might indicate the presence or absence of a particular cladoceran community type. Knowledge of the factors controlling cladoceran communities at different scales is important for understanding the additional factors being incorporated into the analogue matching procedure by the addition of cladoceran data, and

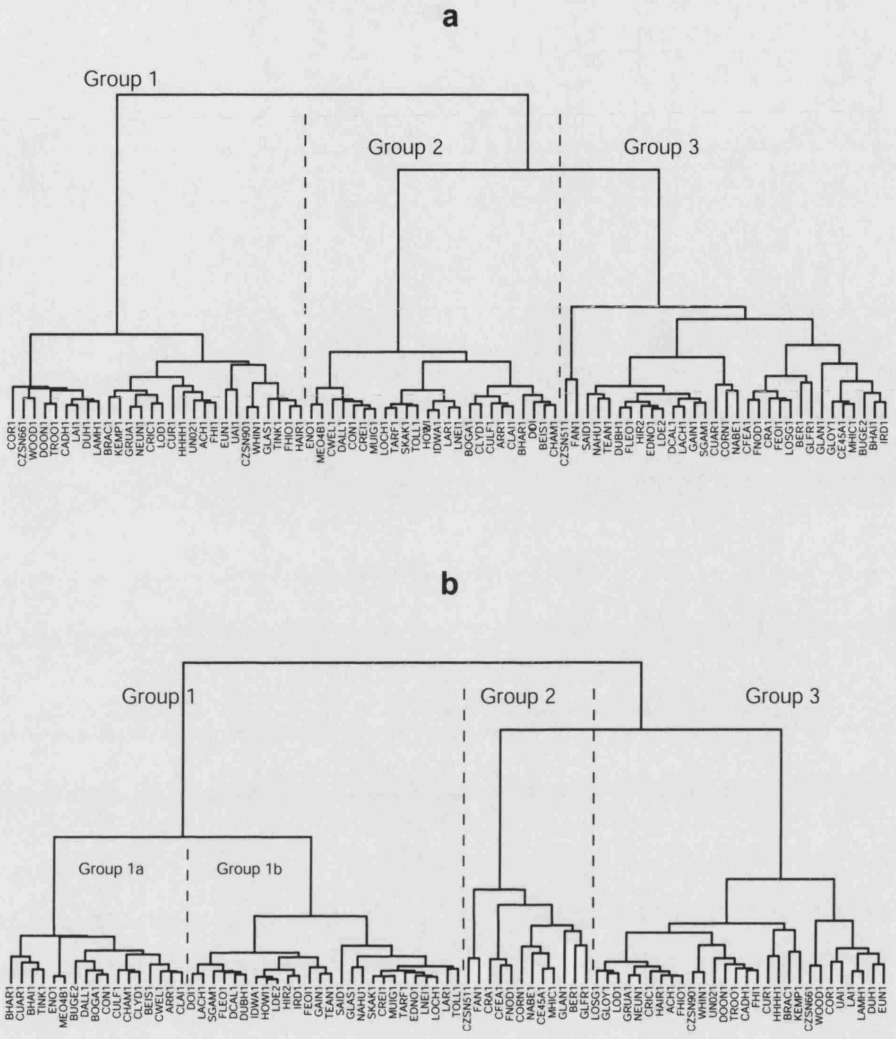


Figure 52: Dendrograms showing the results of a hierarchical cluster analysis using Ward's minimum variance method of a Euclidean distance matrix (a) and a squared chord distance matrix (b) of the cladoceran species abundance data for the 83-lake training set.

to inform palaeolimnological and contemporary studies, thus allowing a fuller interpretation of cladocera species data.

A hierarchical cluster analysis using Ward's minimum variance method was performed on the cladoceran species abundance data. The Euclidean distance is not appropriate for representing species abundance data where those data contain many zeroes because the Euclidean distance takes an absence of a species from both samples being compared as contributing to the similarity of the two samples.

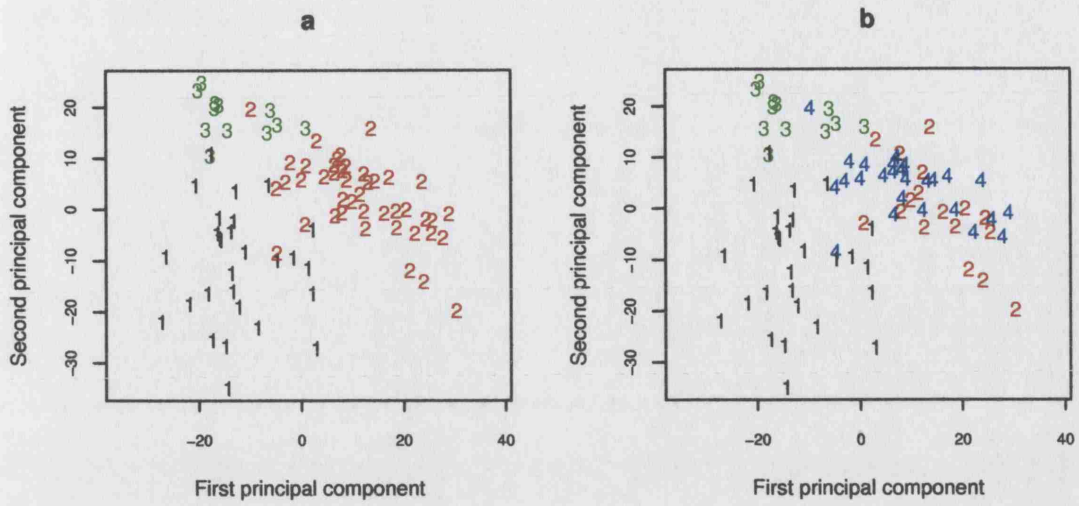


Figure 53: Biplots showing the results of the hierarchical cluster analysis of the squared chord distance matrix of species abundance data for three clusters (a) and four clusters (b) on the first two principal components.

This is not the case where sampling error may mean that a species could be present in a sample but not recorded. Such an error is routinely encountered in aquatic ecology where one is essentially sampling blind; we cannot be certain that the absence of a particular taxon is due to the environmental conditions or random sampling error, and as such extra weight should not be given to species absences when comparing similarity (Legendre and Legendre 1998). This problem is most apparent when sampling long environmental gradients along which replacement of species will occur. In the case of the cladoceran abundance data, only short (< 2 s.d.) gradients are present, but many of the species found are rare, occurring in only a few samples each. As such the species data matrix is quite sparse and the double zero problem might be more pronounced than would normally be expected for these shorter gradients.

The analysis was performed twice; once using a standard Euclidean distance matrix to form the dendrogram and then using a squared chord distance matrix. The two analyses were performed to investigate differences between the two approaches and assess the suitability of using the Euclidean distance for clustering the cladoceran abundance data.

Figure 52 shows the results of two cluster analyses with the upper dendrogram (Figure 52a) based on a Euclidean distance matrix and the lower dendrogram (Figure 52b) on a squared chord distance matrix. The analysis of the Euclidean distance matrix indicates the presence of 3 main clusters of sites of roughly equal size, whilst the dendrogram of the

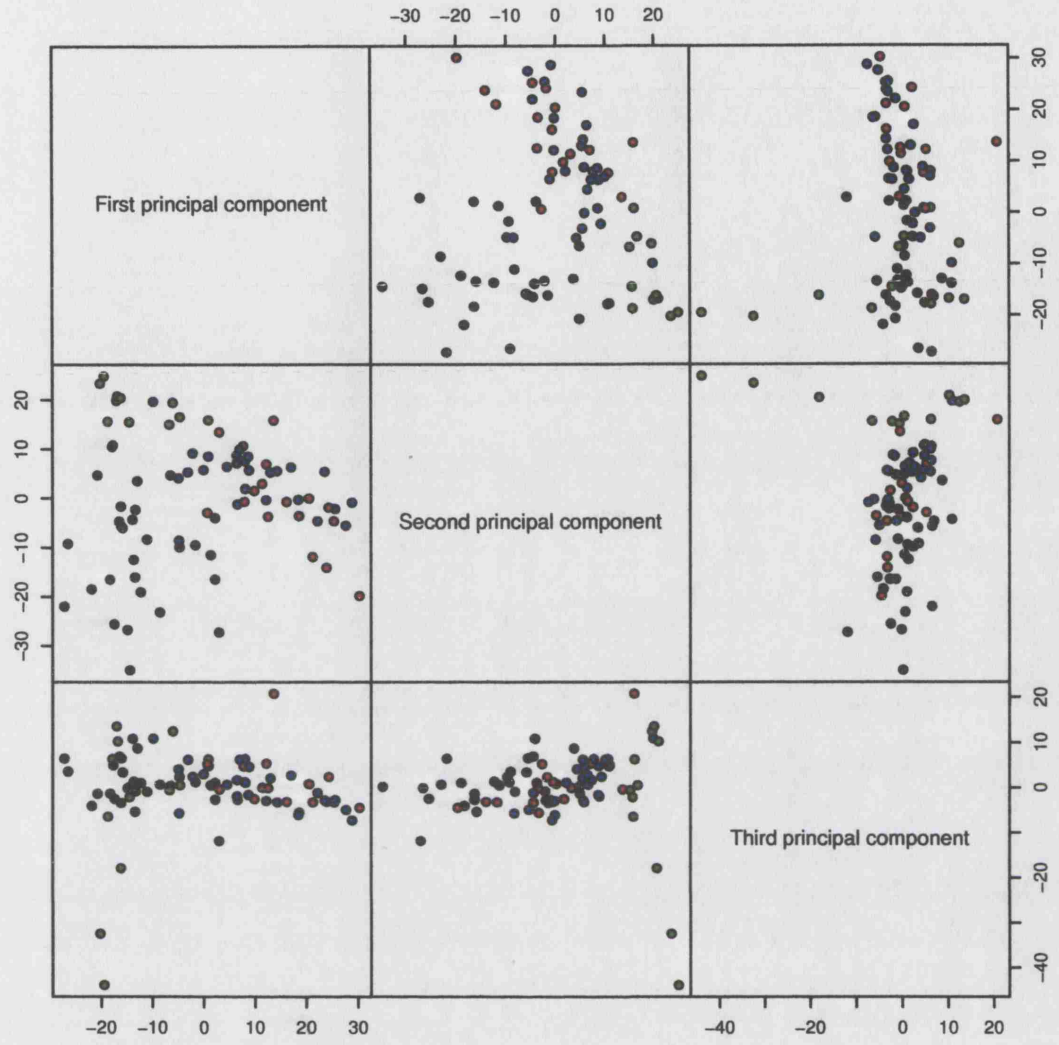


Figure 54: Scatterplot matrix illustrating the positions of the 83 lakes in the first three dimensions of the principal component space of samples. Lakes are classified according to the results of the hierarchical cluster analysis (Ward's clustering method, squared chord distance matrix). Group 1 (black circles), Group 2 (red circles), Group 3 (green circles) and Group 4 (blue circles)

squared chord distance matrix suggests that three or perhaps four clusters of sites exist in the data. By analysing the species abundance data using PCA and plotting the sample scores on the first two principal components according to their cluster membership, a graphical representation of the clustering solution can be achieved. The PCA extracts the main patterns in the data and a measure of the validity of the clusters can be achieved by examining the distribution of the clusters in ordination space.

Figure 53 shows two biplots of the cluster analysis results for both the three and four cluster solutions. Whilst the 3 groups (Figure 53a) are fairly well defined in principal component space with little overlap, the clustering of the 4 groups (Figure 53b) shows a

large degree of overlap between groups 2 and 4. This suggests that the splitting of group 2 into two clusters is not a justified partition given that the lakes within the two clusters are very similar in ordination space. The differences between groups 2 and 4 may not be associated with the first two principal components, this variance instead being expressed on lower axis of the PCA. Figure 54 shows a scatterplot matrix for the first three principal components plotted against each other. The sample locations within the ordination space described by each pair of principal components are coded according to cluster membership. The red and blue circles represent groups 2 and 4 respectively and show that sample locations in ordination space for both groups are quite similar. Only group 3 shows any great distinction from the other points when plotted on the third principal component. Furthermore, a screeplot (not shown) of the variance explained by each axis of the PCA suggests that only the first two principal components explain significant amounts of variation; which is reflected in the level of separation of the samples along this ordination axis in Figure 54. These results suggest the presence of three distinct clusters in the cladoceran abundance data for the cluster analysis of both a Euclidean and a squared chord distance matrix.

In this and the previous chapter the point has been made that hierarchical classifications do not guarantee to find *the* optimal solution to the clustering problem. Hierarchical clustering methods aim to reduce the immense number of configurations that need to be compared to find the optimal configuration to a more manageable size. There is, therefore, a trade off between computational ease and the degree to which the resulting configuration is likely to be the optimal configuration. *k*-means clustering is an iterative partitioning method, whose algorithm does not work on a hierarchical basis. Both Ward's method and *k*-means clustering aim to form clusters that minimise the within group sum of squares (i.e. have greater similarity within group) for the clusters. The measure of similarity in both can be shown to be a Euclidean distance. Therefore, *k*-means clustering can be used to *update* the configuration arrived at via Ward's minimum variance clustering to achieve a lower within group dissimilarity. If the Ward's minimum variance clustering results in a below optimal configuration of samples *k*-means clustering, using the cluster centroids of the Ward's analysis should produce a more optimal solution. Again, there is no guarantee that this configuration will be *the* optimal configuration, but the groups should be more discrete following the *k*-means analysis.

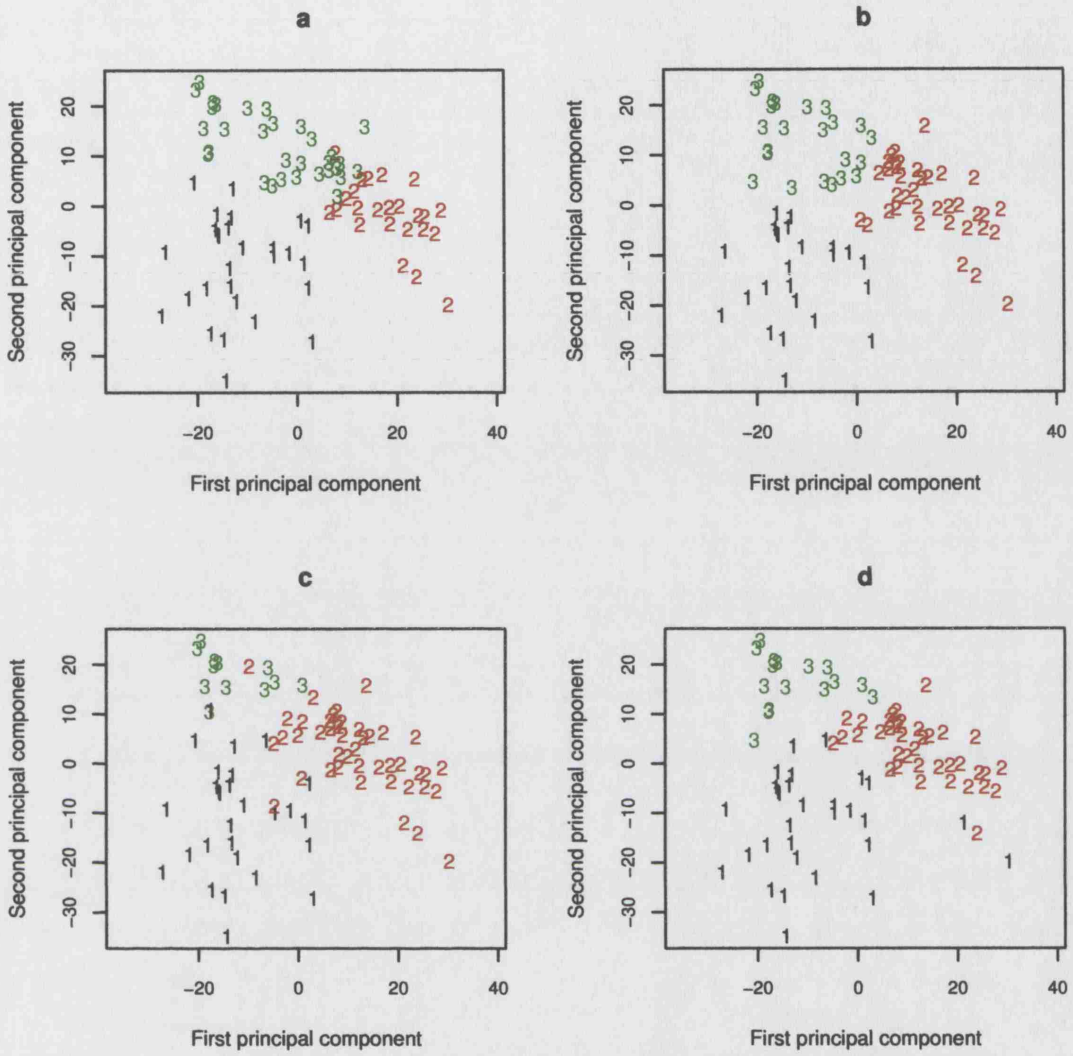


Figure 55: Biplots showing the results of a cluster analysis of a Euclidean distance matrix (a and b) and a (squared) chord distance matrix (c and d) of cladoceran species abundance data. Biplots a and c show the results of a Ward's minimum variance cluster analysis and biplots b and d show the results of the k-means partitioning. The distance measure used in analysis d is in fact the chord distance (obtained by performing k-means partitioning on square root transformed abundance data) not the squared chord distance used in c. Whilst being numerically different the dissimilarity matrix for the chord distance and its squared form should contain the same information.

It can be shown that the chord distance between any two samples is the equivalent of the Euclidean distance between those two samples following a square root transformation. Chord distance and its squared form produce similar results to one another.

The distance measure implicit in *k*-means partitioning is the Euclidean distance. By applying a square root transformation to the species abundance data prior to partitioning using the *k*-means algorithm the distance measure used in the analysis will be the chord distance of original, untransformed abundance data. The chord distance has been shown to be a good compromise between giving equal weight to the rare and common taxa and

reducing the weighting of the rare taxa in the analysis (Overpeck *et al.* 1985). The chord distance also has suitable properties for analysing percentage abundance data as used here.

Figure 55 shows the results of these classifications in principal component space. The configurations resulting from the Ward's minimum variance clustering have been updated following the *k*-means partitioning which suggests some improvement in the configurations over the hierarchical approach. It is clear from the biplots above (Figure 55b and d) that the *k*-means partitioning produces tighter clusters that Ward's method for both dissimilarity measures. However, both clustering techniques produce very similar classifications of the samples for each dissimilarity matrix, with the Euclidean distance measure showing the more discrete clusters when plotted in PCA space. This would suggest that the Euclidean distance is a suitable distance measure for the classification of the cladoceran species abundance data.

However, when the species/samples data matrix is reordered to reflect a clustering of the samples and a clustering of the species (i.e. which species co-occur) it becomes apparent the main differences between clusters are in the relative proportions of *B. coregoni* and *B. longispina* in the samples. One group of sites is dominated by *B. longispina* and one by *B. coregoni*, with a third group having lower proportions of these two taxa and a more chydorid dominated assemblage. This detail is captured most faithfully by the analysis of the squared chord distance matrix using Ward's minimum variance method. The doubly-ordered species abundance matrix and dendrograms of the species and samples cluster analyses for the squared chord distance matrix and Ward's clustering method is shown in Figure 56.

This type of display illustrates the main properties of the cladoceran species abundance data. The immediate feature is the number of rare taxa in the data set with abundances of no greater than 10% though generally abundances are lower than this level (species Group 4, SG4). This highlights issue of sparse data matrices even along short gradients.

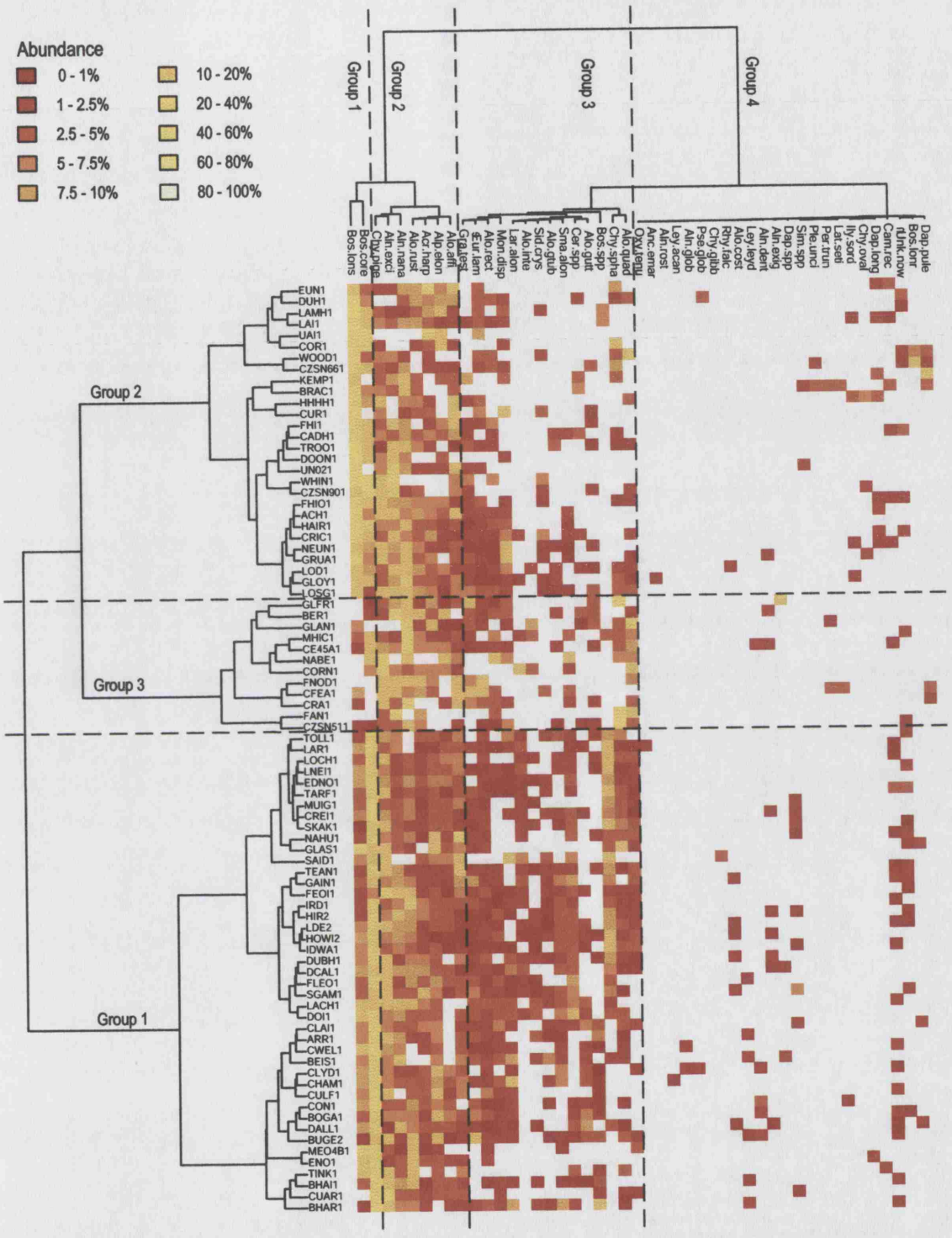


Figure 56: Doubly reordered representation of the species percentage abundance data matrix. The rows of the matrix are ordered according to a Ward's minimum variance cluster analysis and illustrate those lakes in which a similar community of Cladocera is found. The columns of the matrix are ordered on the basis of species associations as determined by a hierarchical cluster analysis (Ward's method, squared chord dissimilarity matrix). Species associations are recurrent groups of species that co-occur across the training set. The species abundances for each sample are represented by blocks of colour. Cool colours (reds) indicate low percentage abundance in a sample, whilst hot colours (yellows) reflect high percentage abundance. White spaces indicate the absence of a particular species from that sample. The legend indicates the class divisions for the species abundance matrix. Greater class division at the low-abundance end of the scale has been applied in an attempt to emphasise subtle differences in those taxa that are relatively rare in the samples and to reduce the dominance of the diagram by the very abundant species (e.g. B. longispina *and* B. coregoni*). The group numbering for the dendrogram on the right of the plot (the classification of lakes) reflects the group numbering used in the other figures in this section.*

The remaining three groups of species, or *species associations*, contain those taxa that are much more abundant across part or all of the data set. Species Group 1 (SG1) contains the two species of *Bosmina* identified to species level (*B. longispina* and *B. coregoni*). These two species show a clear and strong association being found consistently together in the majority samples, usually at high abundances. Species Group 2 (SG2) contains a group of abundant chydorid taxa found throughout the data set. Again abundances are generally high (20-40%), though higher and lower values are routinely present. Species Group 3 (SG3) also contains a range of chydorid species, yet these taxa do not occur across the whole range of the data set. Also, abundances for these taxa are generally lower than those for SG2.

There are three main patterns in the cluster analysis of the lakes based on species composition. Lake Group 3 (LG3) contains samples where *Bosmina* contributes only small abundances to the sample total or is absent from the sample altogether. These 12 lakes are dominated by those species in SG2 with contributions, albeit sparsely and at low levels, from species in SG3. LG3 is quite distinct from the other samples present in the training set.

The second major pattern in the Cladocera species abundance data is the relationship between the proportions of *B. longispina* and *B. coregoni*. Samples in Lake Group 1 (LG1) are consistently dominated by *B. coregoni* whilst those in Lake Group 2 have higher proportions of *B. longispina*. *B. coregoni* is present at high abundances in both lake groups whilst *B. longispina* is only very abundant in LG2, with the species being moderately abundant in LG1. Another interesting pattern in the *Bosmina* is that of the undifferentiated *Bosmina* (*Eubosmina*) spp. (coded Bos.spp in the diagrams) in LG1. With only a few exceptions, only samples in LG1 contained *Bosmina* (*Eubosmina*) spp. These *Bosmina* (*Eubosmina*) spp. bear properties of both *B. longispina* and *B. coregoni*, particularly in the headshields and did not clearly belong to either species group. It is interesting that this species type is mainly restricted to LG1. The other significant difference between LG1 and LG2 is the presence of species from SG3, and to a lesser extent SG4, in the samples. The lakes of LG1 contain a much richer suite of Cladocera than either LG2 or LG3.

Linear discriminant analysis (LDA, also known as canonical variates analysis or CVA) can be used to select a linear combination of variables that best discriminates, or separates, *a priori*

groupings of samples. LDA was used to investigate which of the twenty-four measured physico-chemical variables best discriminates the lake groupings of the minimum variance cluster analysis and the *k*-means partitioning (squared chord distance matrix). LDA was performed using the lda() function in the MASS package (Venables and Ripley 1999) for R on the groups defined from the Ward's minimum variance cluster analysis and the *k*-means partitioning (Euclidean distance matrix). Table 37 shows the amount of between-group variance explained by the first and second linear discriminant functions, whilst Figure 57 shows the positions of each of the 83 lakes along these two linear discriminant functions coded by group number. Only g-1 (where g = the number of *a priori* defined groups) linear discriminant functions can be extracted so for the 3 groups of lakes identified above two linear discriminant functions are sufficient to explain fully the between-group variance.

The two clustering methods result in very similar configurations of samples, so it is not surprising that the 24 measured physico-chemical variables explain similar amounts of between-group variance on both the first and second linear discriminant functions. As discussed above, the *k*-means partitioning method (Euclidean distance matrix) led to the tightest clustering of the 83-lakes in principal component space. Almost 80% of the *k*-means-defined between-group variance is explained by the first linear discriminant. The coefficients of the linear discriminant functions, also known as canonical variates, indicate that the first linear discriminant function (LD1) is most closely correlated with catchment area (2.028), lake area (-1.492) and the ratio of lake area to catchment area (1.424). Other variables that are strongly correlated with LD1 are chloride (1.216), magnesium (-1.009) and total organic carbon (0.865).

Table 37: Summary results of the LDA of the Ward's minimum variance and the k-means cluster analyses.

Cluster method	**Between-group variance explained by:**	
	Linear discriminant 1	**Linear discriminant 2**
Ward's Minimum Variance Clustering	0.7097	0.2903
k-means partitioning	0.738	0,262

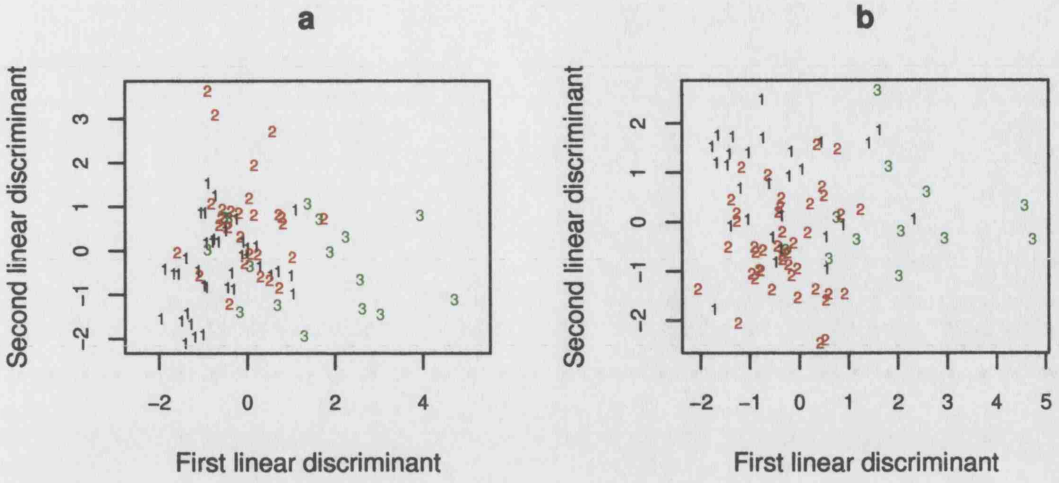


Figure 57: Linear discriminant analysis biplots showing the result of the LDA of the Cladocera lake clusters identified using Ward's minimum variance cluster analysis (a) and k-means cluster analysis. Lake clusters are identified by number, see text for details.

Figure 58 shows the histograms of the probability density function for each of the three defined groups on LD1. The position zero on the x-axis is the discriminant index, and the dashed lines indicate the mean position along the discriminant function for each respective group. Whilst there is still some degree of overlap between the groups it is clear that Groups 1 and 2 are somewhat different to Group 3 along LD1. The means for Group 1 and Group 2 lie on the negative side of the discriminant index whilst the mean for Group 3 lies much further to the right on the positive side of the discriminant index.

Groups 1 and 2 are associated with larger lakes and those lakes that occupy a greater proportion of the total catchment area, whilst Group 3 lakes are more associated with smaller lakes with lower chloride and TOC concentrations. Group 3 lakes are characterised by lower proportions of planktonic *Bosmina* species and a much richer chydorid community than Groups 1 and 2.

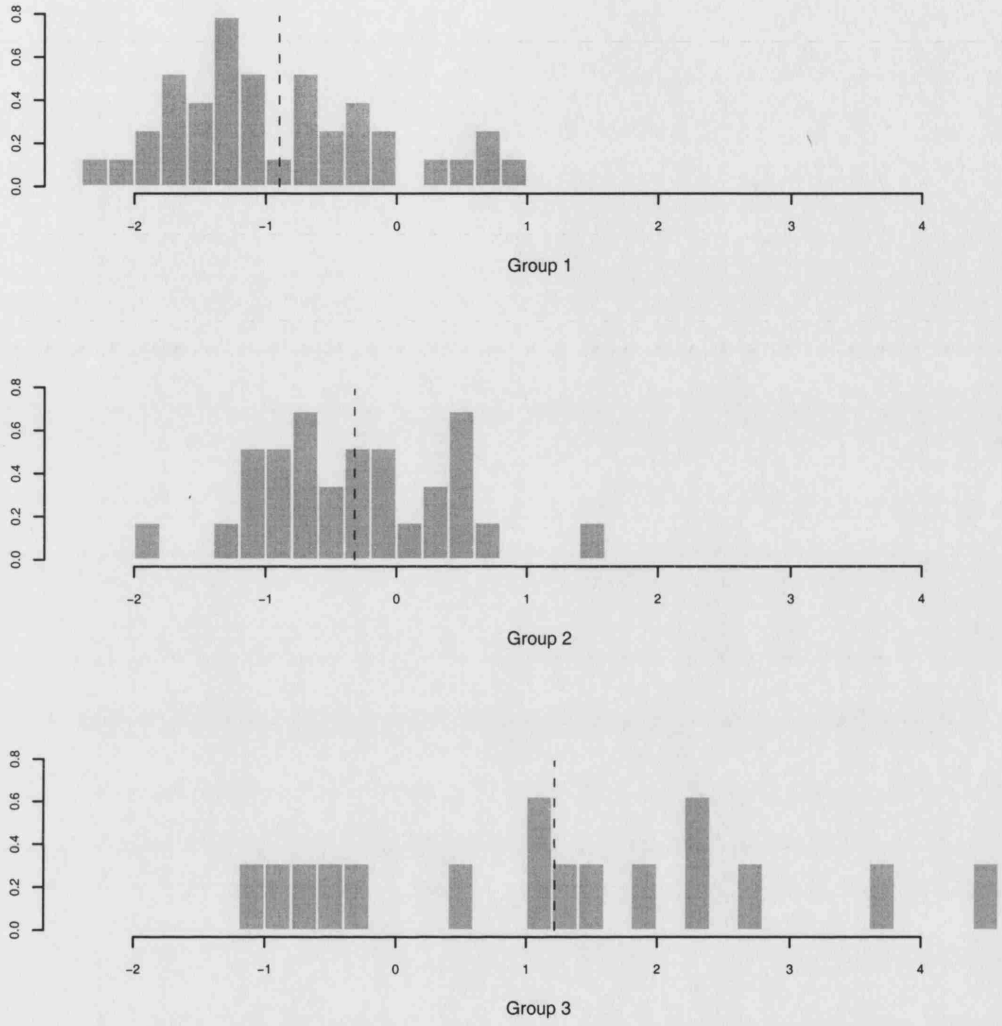


Figure 58: Histograms showing the probability density function of the positions on the first linear discriminant axis for each of the a priori *defined groups.*

Smaller lakes are more likely to have a larger proportion of their area in the littoral zone and as such are likely to have a greater proportion of the total area available for exploitation by chydorids. Larger lakes, with smaller proportions of their area in the littoral zone will conversely have less available habitat to support a diverse chydorid population and planktonic taxa should be more dominant. This pattern is reflected in the results of the LDA, with the chydorid-rich Group 3 being associated with smaller lakes. It is interesting to note that maximum lake depth is not as strongly associated with LD1 (-0.575) as one might expect given the above hypothesis.

One might expect that chydorid-rich sites might also be correlated with shallower lakes with a larger proportion of the available habitat located in the littoral zone. The planktonic

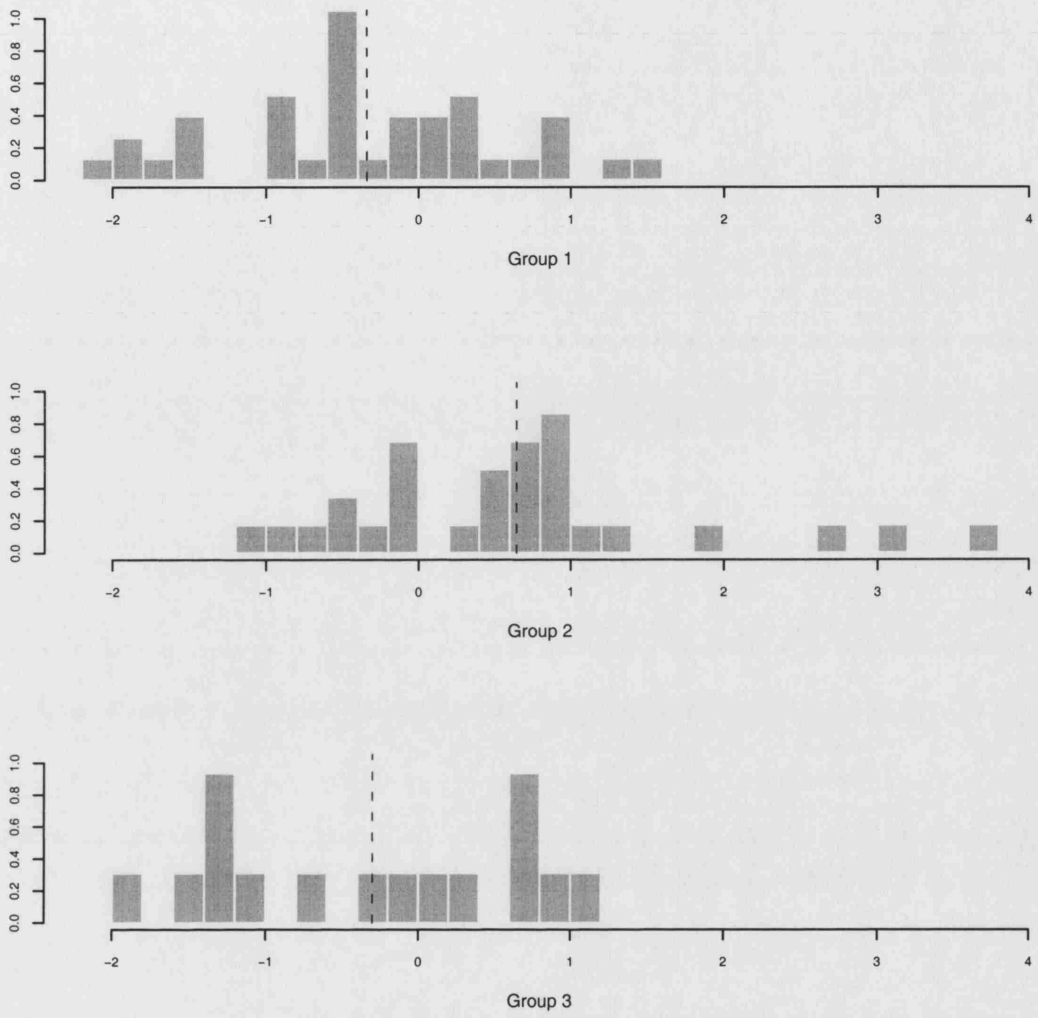


Figure 59: Histograms showing the probability density function of the positions on the second linear discriminant axis for each of the a priori *defined groups.*

Cladocera-dominated systems are associated with the deeper sites but maximum lake depth is not as good at discriminating between the three groups of lakes as the other variables mentioned above.

Catchment area and the lake area to catchment area ratio contribute strongly to the second linear discriminant function (LD2). A range of hydrochemical parameters also contribute to the between-group variance on the LD2; conductivity (-1.057), calcium (-1.290), potassium (-0.958) on the negative side of the discriminant index and chloride (1.551) and sulphate (0.964) on the positive side. Maximum lake depth also contributes quite strongly to the discrimination between groups on LD2 (-0.769).

Figure 59 shows the probability density function histogram of the canonical variates for the second linear discriminant function. Using the linear coefficients of the variables for LD2 Figure 59 suggests that Groups 1 and 2 contain sites with higher ionic strengths and in larger catchments than the lakes in Group 2. Group 2 lakes are shallower and whilst being of lower ionic strength than either of the other two Groups, tend to have a larger contribution to the total ionic strength contributed by chloride inputs.

4.5 Distribution of selected cladoceran taxa in upland areas of the UK

The following section of the thesis presents a series of maps that illustrate the distribution of the main species of *Cladocera* in the upland fresh water systems in the 83-lake training set (Figure 60 to Figure 65).

The main feature of these maps is that they illustrate the ubiquity of the *Cladocera* as a group. With the exception of *R. falcata* and a couple of other rare taxa in the training set (*B. longirostris* and *A. costata*) all the remaining taxa illustrated in Figure 60 through Figure 651 occur throughout the acid sensitive upland areas of the UK.

Figure 60 shows the distributions in the training set of 4 chydorid taxa; *A. harpae*, *A. nana*, *A. affinis*, and *A. excisa*. All four taxa are common components of acid sensitive upland systems forming between 25 and 50% of the cladoceran community in many systems. As with the other chydorid taxa presented in these maps, none of the four chydorids dominates any given sample. This is in stark contrast to the dominance of the two acid tolerant bosminids, *B. coregoni* and *B. longispina*, which regularly contribute over 50% of the cladoceran sub-fossil community.

The distribution of *A. harpae* complements the results of the GLR (Figure 49a), which showed that whilst *A. harpae* was present in lakes across a wide range of altitudes, it was consistently present in the highest altitude lakes in the training set. Figure 60a shows high abundances for this taxon in the Snowdonia region of Wales, the Cairngorm region of Scotland and the high altitude corrie lochs of northwest Scotland. *A. nana*, *A. affinis* and *A. excisa* are very common species in acid, oligotrophic waters in the UK and are found at high abundances throughout the training set (Figure 60b).

The distributions of *A. costata*, *A. guttata* var. *tuberculata*, *A. guttata* and *A. intermedia* are shown in Figure 61. *A. costata* occurs rarely in surface waters in the UK, which is illustrated by the presence of the taxon in only 5 of the 83 lakes in the training set. *A. rustica*, a closely related species, is much more common amongst the *Chydoridae* found in the UK (See Figure 62b). The remaining three taxa shown in Figure 61 are commonly found in oligotrophic waters in the UK as illustrated by their presence throughout upland areas of Wales, Galloway and northwest Scotland. The species are not as ubiquitous as the taxa shown in Figure 60, being largely absent from those training set lakes in areas of central Scotland, such as the Trossachs, the Grampians and the Cairngorm plateau.

Figure 62 shows the distributions of *A. quadrangularis*, *A. rectangula*, *A. rustica* and *A. elongata*. The distribution map for *A. quadrangularis* shows that this species has a more restricted distribution than many of the other chydorids in the training set. The species is absent from many of the lochs from northwest Scotland in the training set Figure 60 shows that *A. quadrangularis* is found in lakes with higher than average pH and calcium and TOC levels, and there are less lakes of this type in the training set as a whole. The distribution patterns for the other three taxa shown in Figure 62 are similar to those shown in Figure 60 and contrast with those shown in Figure 61 with *A. rectangula*, *A. rustica*, and *A. elongata* being found in all regions of the UK sampled in the training set.

The distribution of *A. rustica* (Figure 62b) demonstrates that this particular taxon is a common member of acid sensitive, upland lakes in the UK. The patterns of *A. costata* (Figure 61a) and *A. rustica* (Figure 62b) reflect previous observations that, unlike many other areas of Northern Europe and North America, *A. costata* is largely absent from UK surface waters. *A. rustica*, a closely related species to *A. costata*, is known to be much more common in net samples from UK lakes and this is clearly reflected in the sediment samples in the training set (Frey 1964; Fryer 1993).

Analysis of the cladoceran data from the 83-lake training set has shown that *A. elongata* is much more abundant in those lakes at found at higher altitude, with lower pH and a higher sensitivity to acid inputs (See Figure 42 and Figure 51b for example). This preference for higher altitude sites is reflect in the distribution map for *A. elongata* shown in Figure 62d. The largest circles, indicating higher abundances, are found on the high altitude Cairngorm plateau, and the corrie lochans in northwest Scotland.

Figure 63 shows the distribution maps for three planktonic cladocerans, *B. coregoni*, *B. longispina* and *B. longirostris*, and relatively less abundant chydorid taxon, *Camptocercus rectirostris*. An immediately apparent feature of Figure 63 is the contrast between the distributions of *B. coregoni* (Figure 63a) and *B. longispina* (Figure 63c) and that of *B. longirostris* (Figure 63b). *B. longirostris* is largely restricted to more nutrient rich lakes in the UK with higher pH; lakes that are not well represented in the training set. It is quite clear from the distribution map that *B. longirostris* is not a common member of the cladoceran fauna of acid sensitive, upland lakes in the UK.

Conversely, *B. coregoni* and *B. longispina* are ubiquitous members of the cladoceran fauna found in acid sensitive lakes. As mentioned above, these two taxa often dominate that cladoceran community of many of the lakes in the training set and are often found contributing to more than 50% of the cladoceran communities in these lakes. There is a subtle difference in the distributions of the two taxa however. *B. coregoni* is the more abundant taxon in surface waters from northwest Scotland and Wales, whilst *B. longispina* is more abundant than *B. coregoni* in lochs in Galloway, the Trossachs and the Grampians. It is unclear why this is the case, but the results of the cluster analysis of the cladoceran data and the subsequent linear discriminant analysis suggest that, at least from a community level, it is lake water chemistry, particularly conductivity and the contribution of chloride ions to the total ionic strength, that might play a part in this distributional pattern. At the species level, as illustrated in the redundancy analysis presented in Figure 42 and associated text, physical properties appear to account for the differences in the distributions of *B. coregoni* and *B. longispina*. *B. coregoni* is more associated with the high altitude, low pH lakes in the training set, and at average conductivity levels, whilst *B. longispina* is most abundant in the larger, deeper lakes where conductivity levels are lower than average for the training set. *B. longispina* is also more abundant in those lakes with average pH but where levels of aluminium are higher than those where *B. coregoni* is more abundant (c.f. Figure 42).

Camptocercus rectirostris, despite being a relatively uncommon cladoceran throughout the UK, is present in the surficial sediments of acid sensitive lakes throughout the UK (Figure 63d).

Chydorus piger is another common cladoceran, often associated with oligotrophic lakes in the UK. This is reflected in the distribution map for this taxon across the training set lakes. *C. piger* is particularly abundant in the lochs sampled from northwest Scotland (Figure 64a),

and area where sea-salt episodes often affect surface waters in the region. This pattern complements the findings of the RDA shown in Figure 42, which indicates that *C. piger* is associated most closely with those lakes that have a higher conductivity and higher levels of the major ions.

Another member of the genus *Chydorus*, *C. sphaericus*, is also commonly found in upland lakes in the UK, although this species is perhaps better described as a species complex, with different sub-species or varieties likely to be more suited to particular lake conditions rather than notion that *C. sphaericus* as a whole is an taxon to lake conditions. *C. sphaericus* is often found in highly eutrophic lakes in low lying areas of the UK as well as in many acid surface waters. Whilst the distribution map (Figure 64b) largely reflects the cosmopolitanism of this species, it is clear that it is a much less important component of the cladoceran fauna in UK acid sensitive lakes.

The distribution maps for *Eurycercus lamellatus* and *Graptoleberis testudinaria* are shown in Figure 64c and Figure 64d respectively. Both species are found throughout the training set across the range of lakes.

The distribution map for *M. dispar* shows that this taxon, whilst being found throughout the UK, is most abundant in lakes in the north and northwest of Scotland (Figure 65a) and reflects the results of the RDA (see Figure 42) which indicate that this species is associated with the higher conductivity lakes in the training set.

Rhynchotalona falcata is a rare cladoceran generally found in low abundances in northern areas of the UK, and rarely in England (Fryer, 1993). This feature of the distribution of *R. falcata* is illustrated in the pattern of abundance in this taxon across the training set (Figure 65b). *R. falcata* was found in only seven of the 83 sampled lakes, and in each case the lake was from Scotland. This species was not found in any of the Welsh lakes sampled in the training set. Whilst no lakes from England were included in the training set, the absence of the taxon from lakes outside Scotland and the paucity of occurrences throughout the training set conform to the notion that this species is rare in the UK.

The distribution map of the abundance of *Sida crystallina* throughout the training set largely parallels the distributions of the other large-bodies and/or macrophyte associated taxa also recorded (Figure 65c). Like *E. lamellatus* and *G. testudinaria*, *S. crystallina* is found in the majority of lakes sampled in the training set and from across the whole geographical range. These taxa rarely occur in abundances greater than ten percent of the total cladoceran population in a given lake.

The remaining maps in this section (Figure 66 - Figure 69) show the distributions of a number of taxa that are considered rare in this training set (see Table 19). *A. exigua*, for example is found in only 3 of the 83 samples in the training set (Loch na Beiste [BEIS], Loch Dubh Camas an Lochain [DCAL] and Llyn Glasfryn [GLFR]. In the first two samples, *A. exigua* is present in very low abundances (1.25 and 0.55 % respectively), whist it makes up 12.15% of the sample from Llyn Glasfryn. It is unclear why this taxon is particularly abundant in Llyn Glasfryn – but this site is not acid sensitive (alkalinity 413.59 μ eq l^{-1}) and has an almost circumneutral pH (6.9), which might suggest that this species in not tolerant of acid conditions or is just not found in acid waters. It is not clear why this *A. exigua* is not present in other similarly circumneutral lakes in the training set (e.g., WOOD) and the patchy presence of this taxon may be more strongly related to habitat- or predation-related factors and not hydrochemistry *per se*.

A. dentifera is found in 9 samples in the training set but is not very abundant in any of these samples (maximum abundance 1.31%). *A. globosus* and *A. rostrata* are found in a single sample each. A. rostrata is more commonly found in more nutrient rich waters where open littoral-benthic habitats are available (Fryer 1993).

The remaining taxa in this section are extremely rare in the training set, being found in only one or two samples each. It is difficult, therefore, to deduce anything about the distributions of these species from the available data.

A number of other taxa have been identified in the surface sediment samples from the training set, but for which maps of their distribution have not been produced. These taxa are mainly the planktonic *Daphnia* (*D. longispina* and *D. pulex* groups), *Ceriodaphnia* spp. and *Simocephalus* spp. These taxa are not well preserved in lake sediments and it is difficult to identify remains to species level when they do occur. As such it would be inappropriate to present data regarding the distribution of these taxa in acid upland systems.

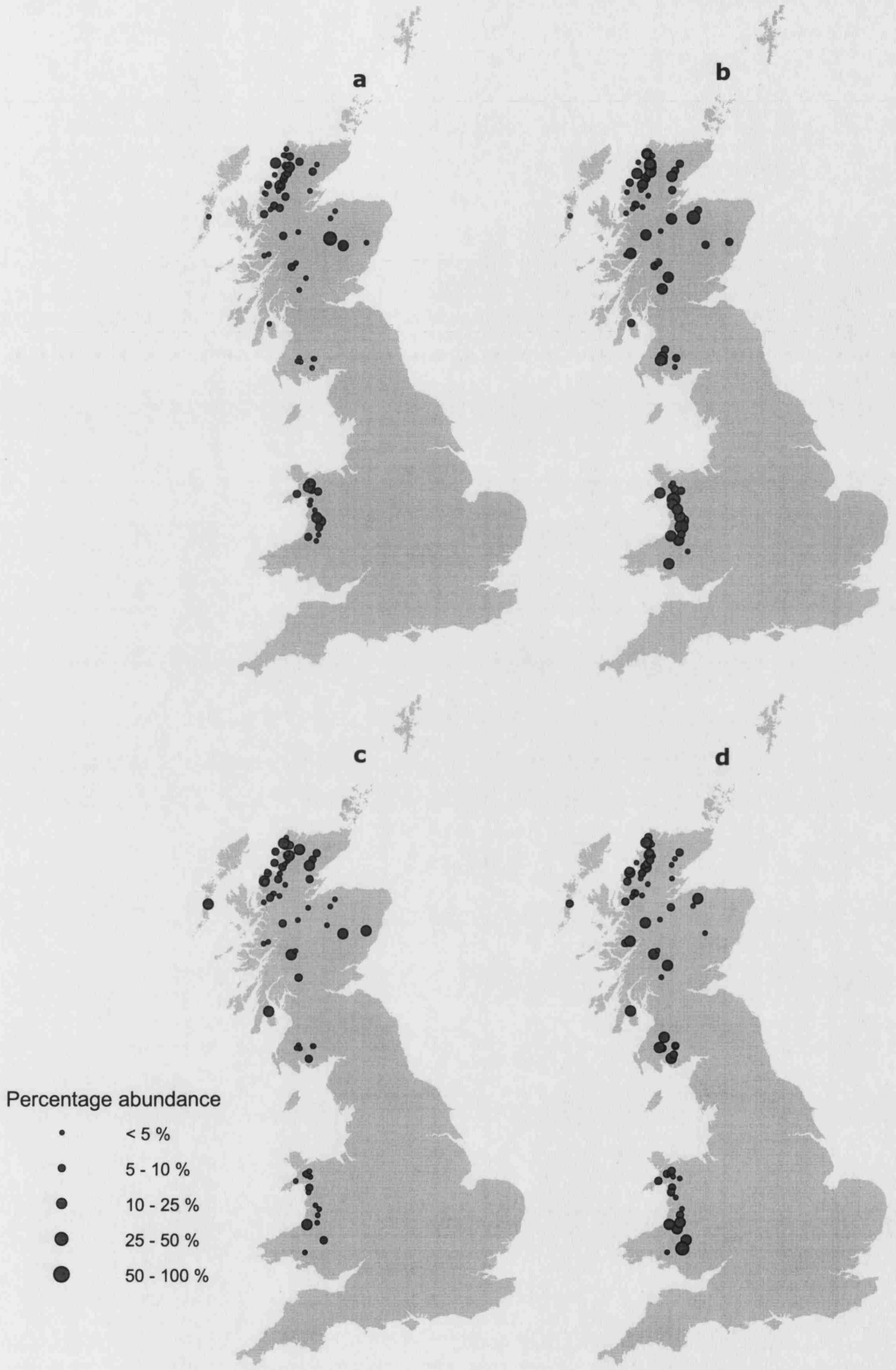


Figure 60: Maps showing the distribution and percentage abundance of A. harpae *(a)* A. nana *(b)* A. affinis *(c), and* A. excisa *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

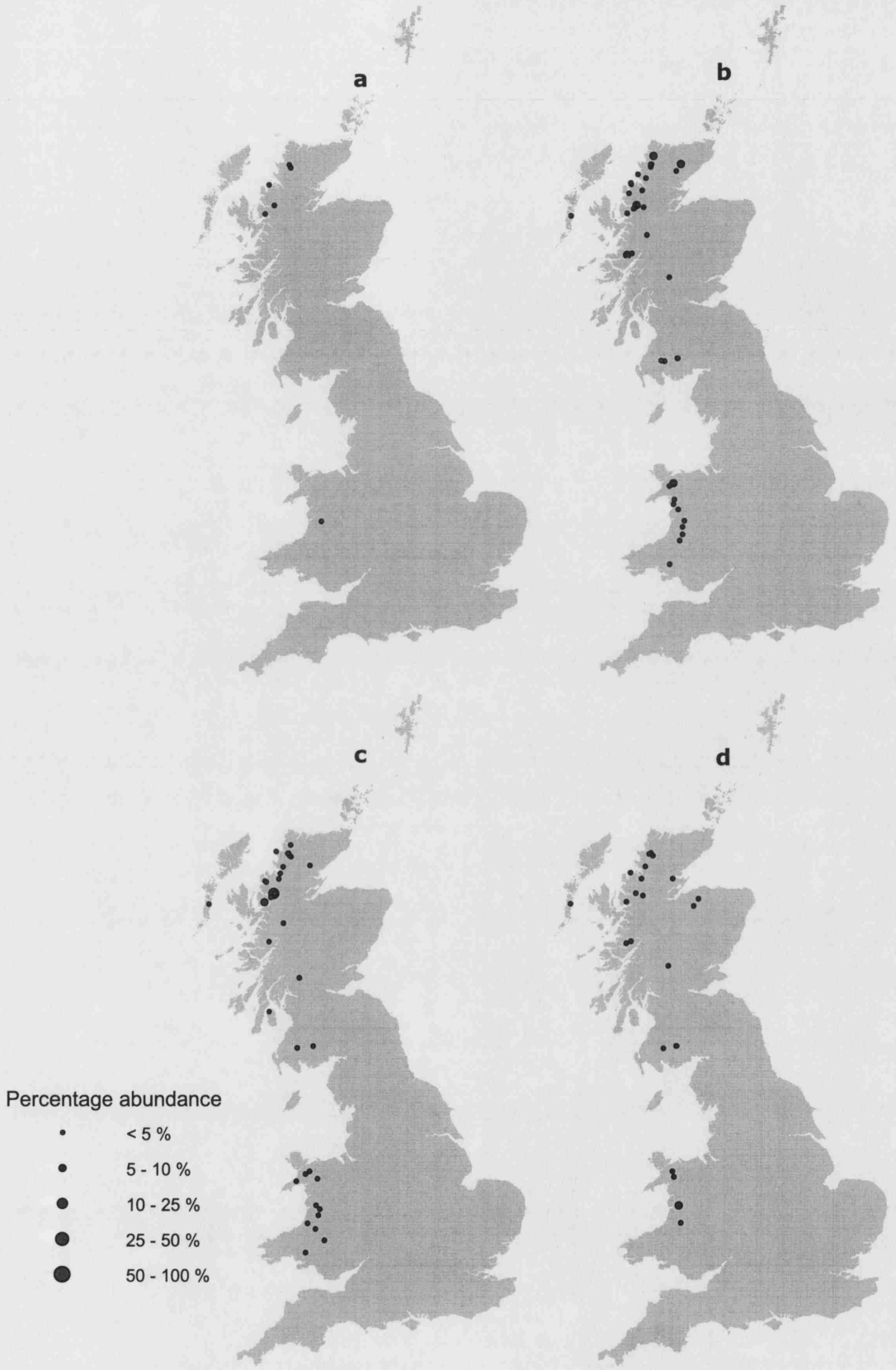


Figure 61: Maps showing the distribution and percentage abundance of A. costata *(a)* A. guttata *var.* tuberculata *(b)* A. guttata *(c), and* A. intermedia *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

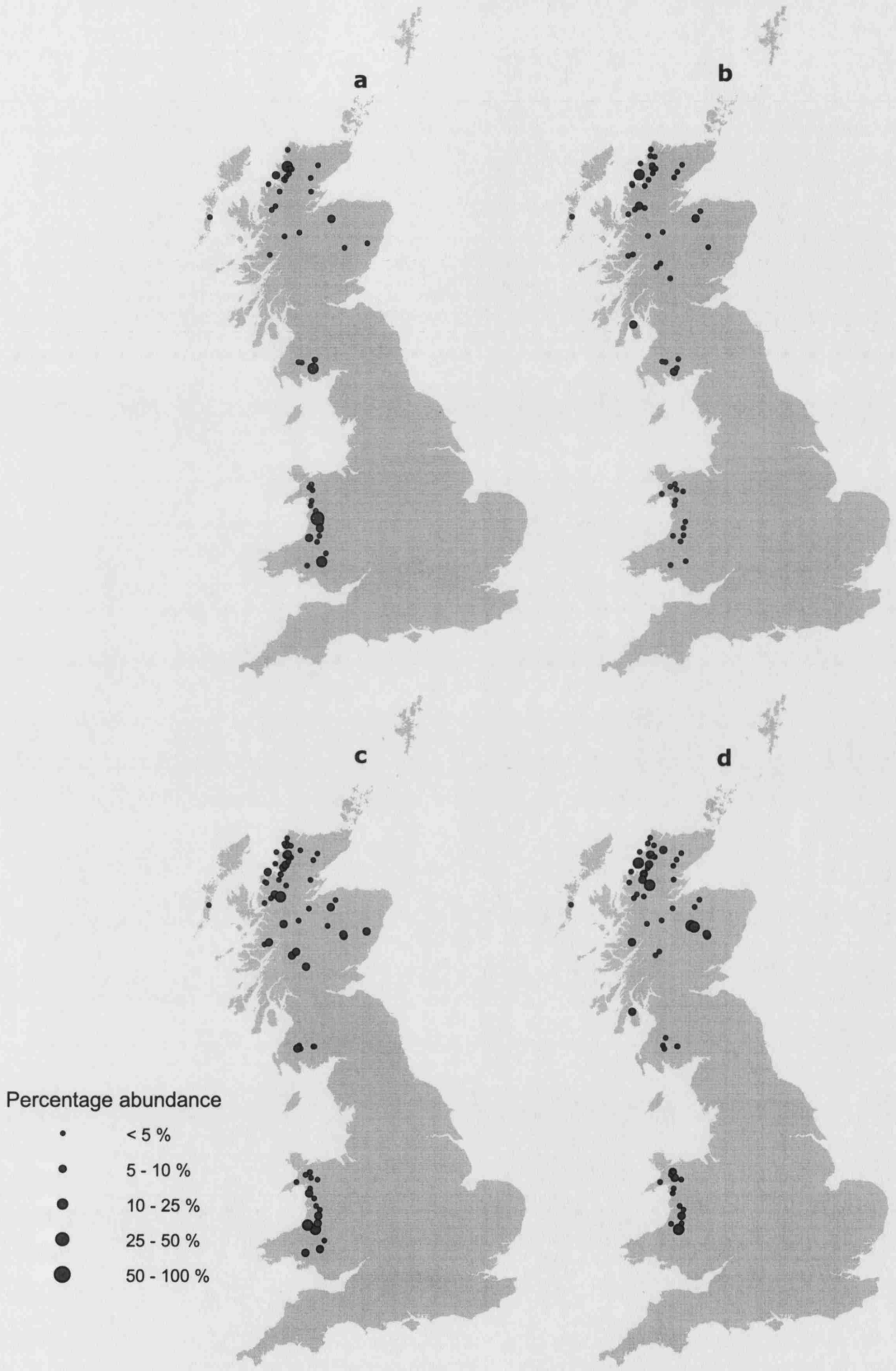


Figure 62: Maps showing the distribution and percentage abundance of A. quadrangularis *(a)* A. rectangula *(b)* A. rustica *(c), and* A. elongata *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

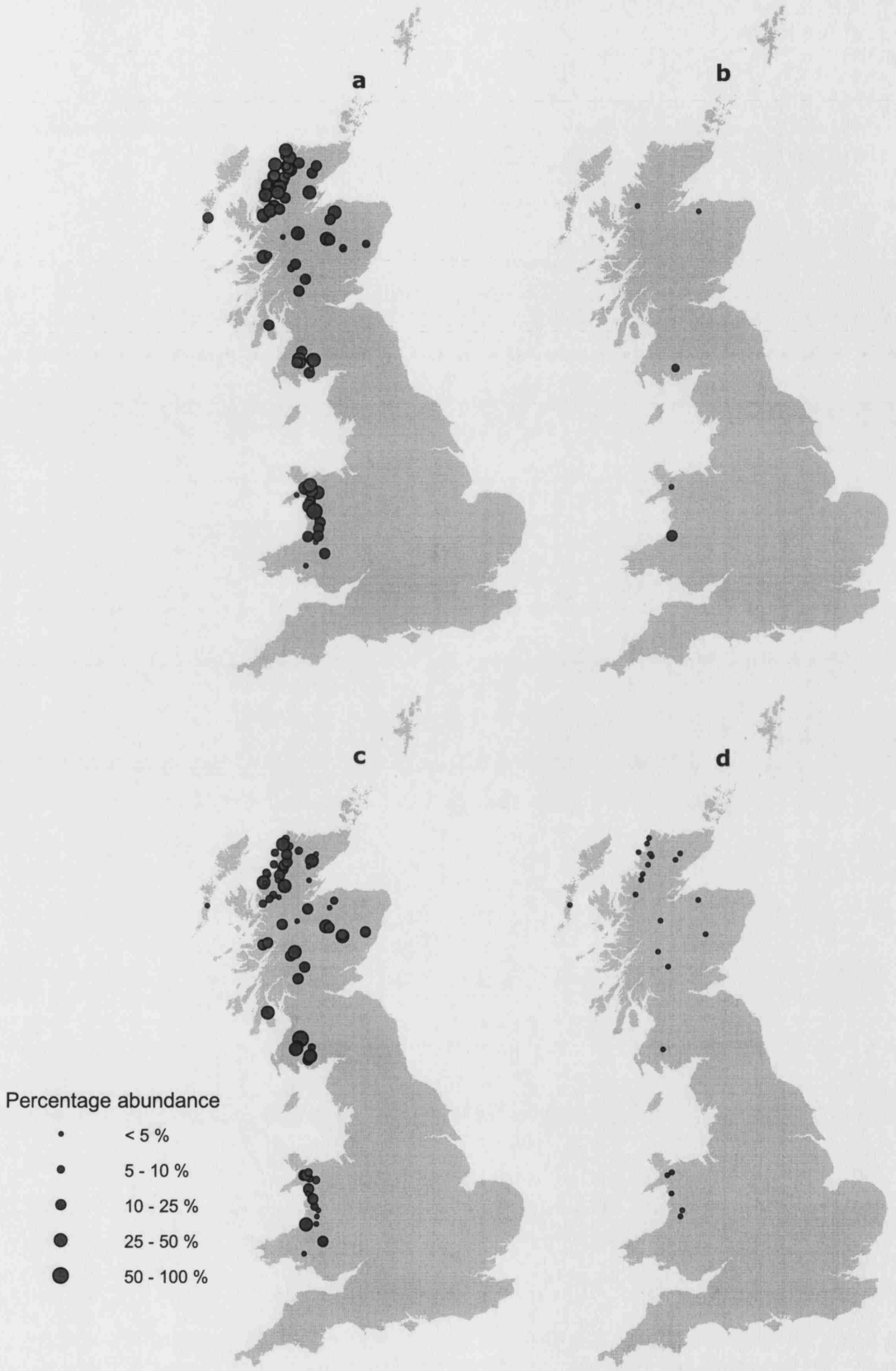


Figure 63: Maps showing the distribution and percentage abundance of B. coregoni *(a)* B. longirostris *(b)* B. longispina *(c), and* Camptocercus rectirostris *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

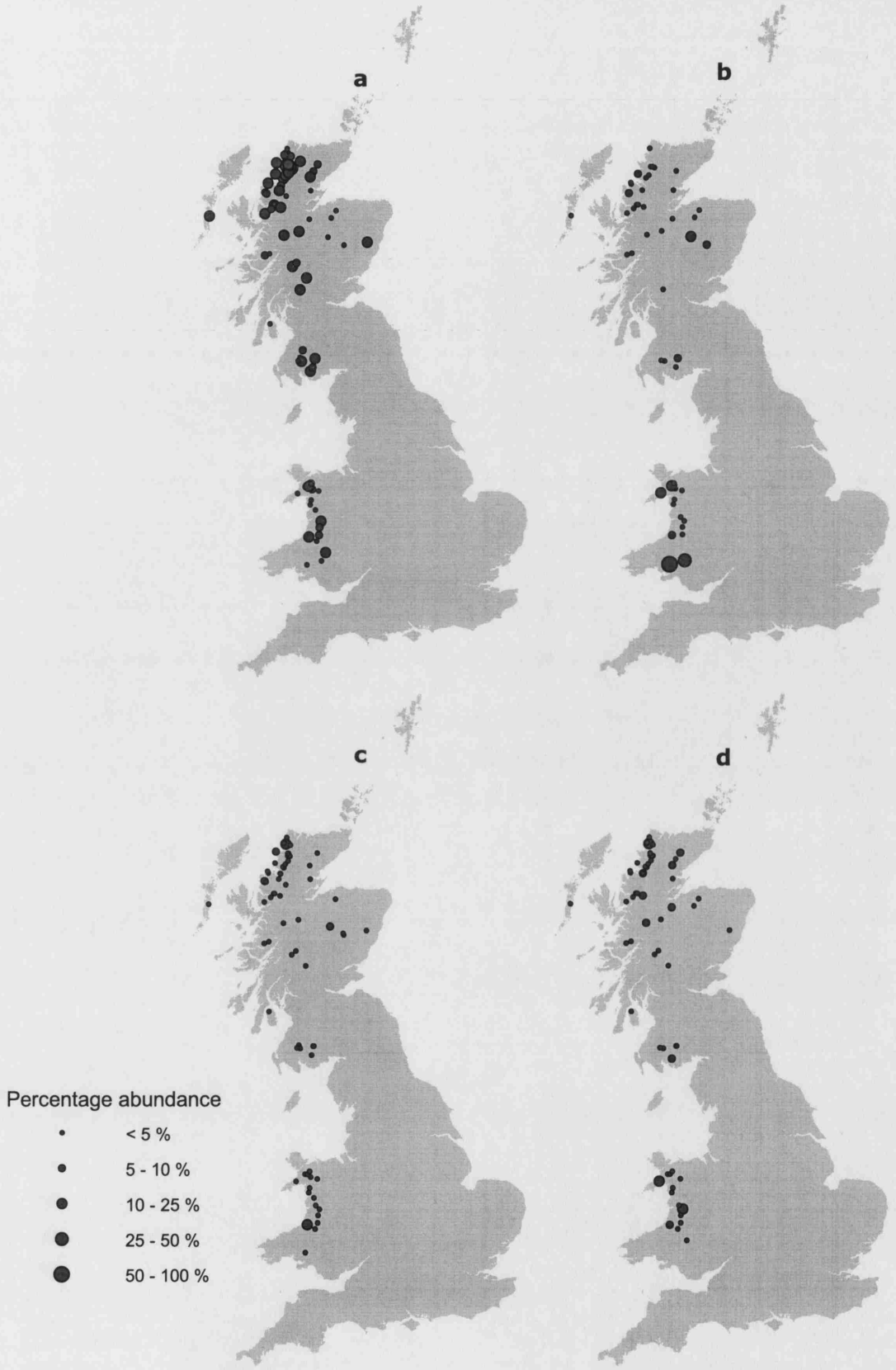


Figure 64: Maps showing the distribution and percentage abundance of C. piger *(a)* C. sphaericus *(b)* E. lamellatus *(c), and* G. testudinaria *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

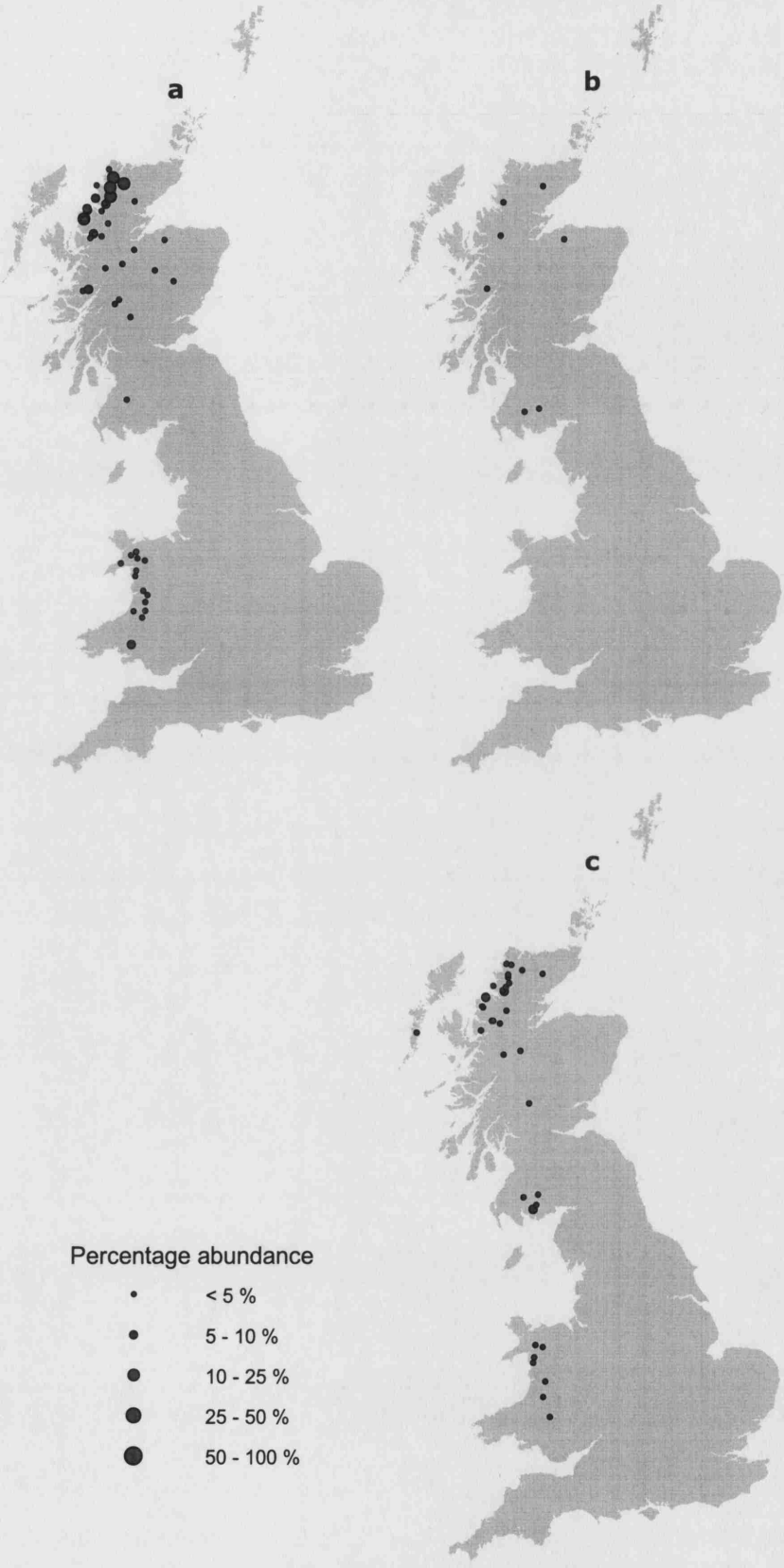


Figure 65: Maps showing the distribution and percentage abundance of M. dispar *(a)* R. falcata *(b) and* S. crystallina *(c) in the 83 upland lakes sampled as part of the cladoceran training set.*

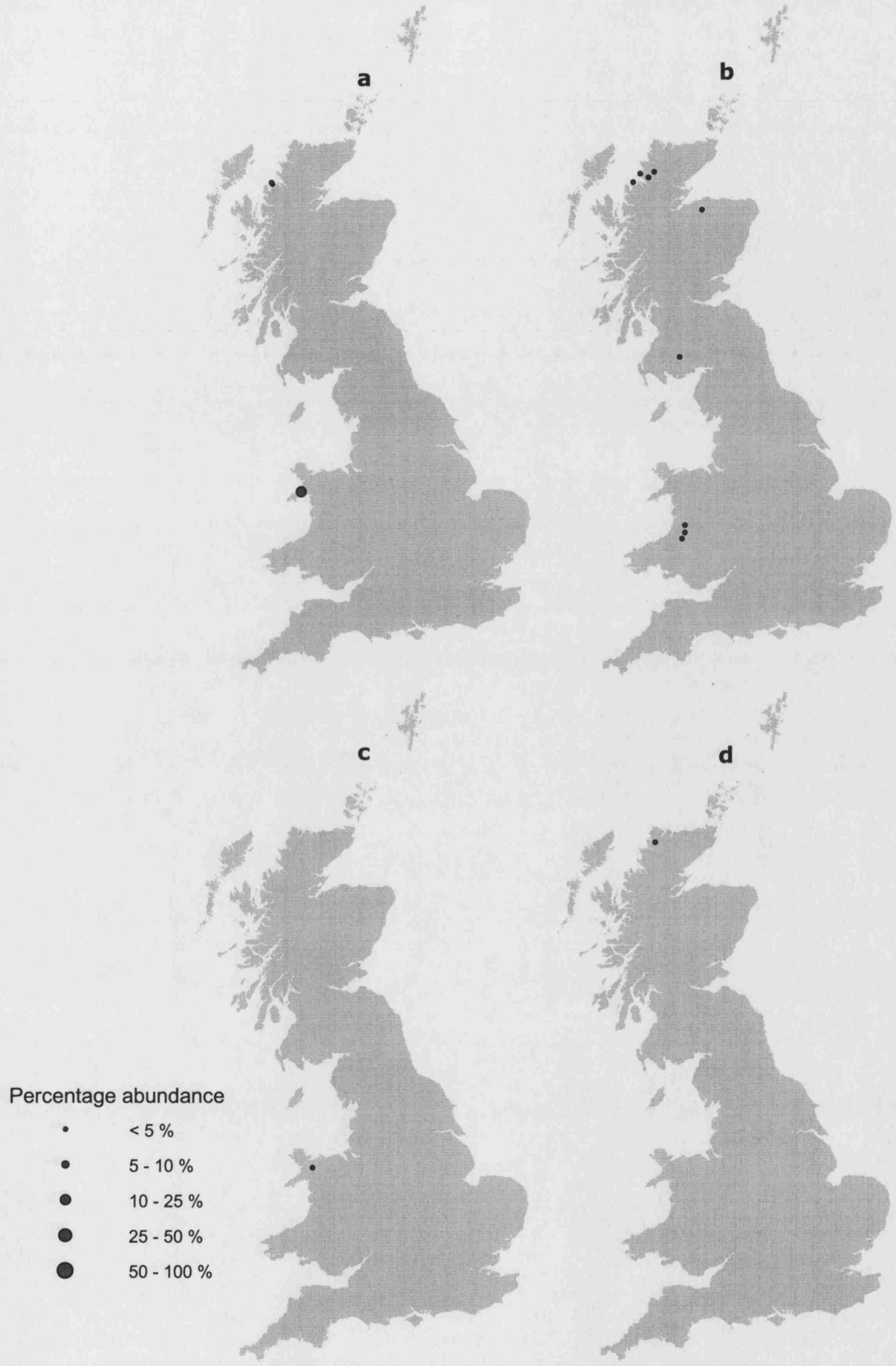


Figure 66: Maps showing the distribution and percentage abundance of A. exigua *(a)* A. dentifera *(b)* A. globosus *(c), and* A. rostrata *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

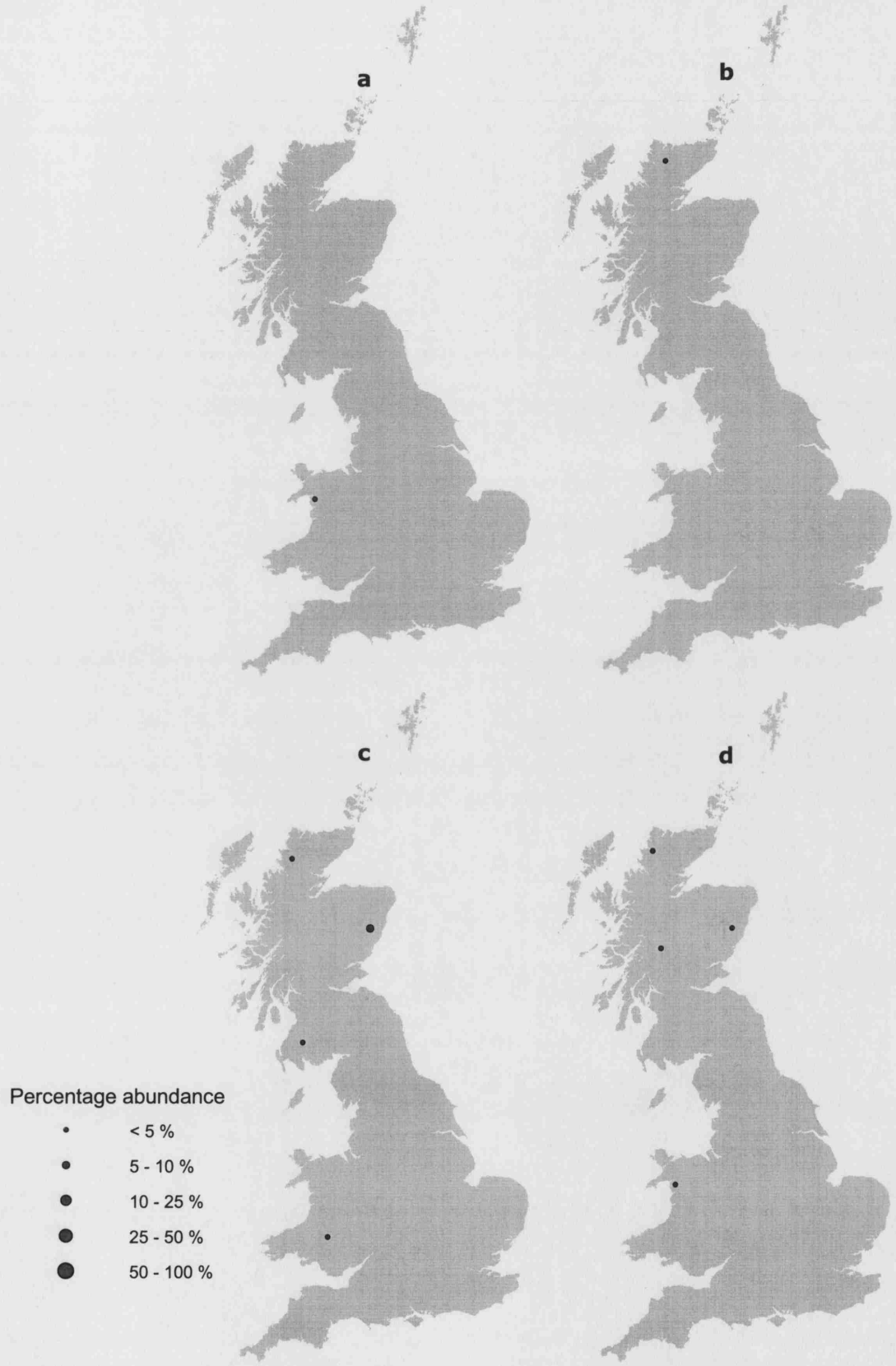


Figure 67: Maps showing the distribution and percentage abundance of A. emarginatus *(a)* C. gibbus *(b)* C. ovalis *(c), and* I. sordidus *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

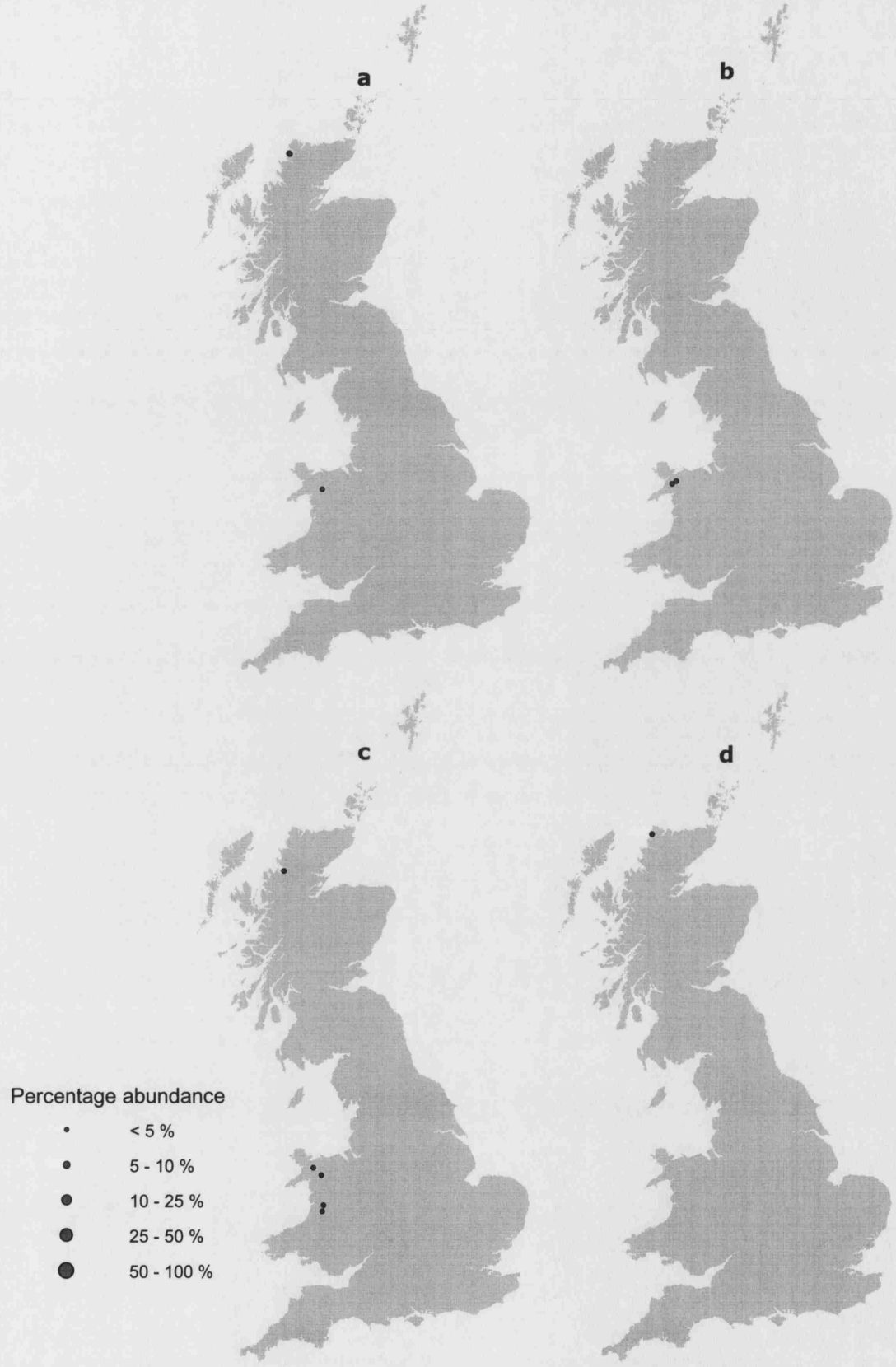


Figure 68: Maps showing the distribution and percentage abundance of L. setifera *(a)* L. acanthocercoides *(b)* L. leydigi *(c), and* O. tenuicaudis *(d) in the 83 upland lakes sampled as part of the cladoceran training set.*

Figure 69: Maps showing the distribution and percentage abundance of P. truncata *(a)* P. uncinatus *(b), and* P. globosus *(c) in the 83 upland lakes sampled as part of the cladoceran training set.*

4.6 Discussion

The primary aim of this thesis is to develop the original analogue matching approach to address some of the problems with diatom-based procedure of Flower *et al.* (1997). In Chapter 3 a 163-sample diatom-based training set was created (the UKAWDDS) to determine whether simply increasing the size of the training set would improve the quality of the matches. A second development of the approach was the inclusion of a second proxy into the matching procedure. The Cladocera were chosen as the second proxy for a number of reasons: the Cladocera leave abundant remains in lake sediments, which are identifiable to species level and below, they occupy a level above the primary producers in the trophic structure of lakes, and the distributions of cladoceran taxa are influenced by a wide range of physical and hydrochemical factors not just those related to acidity.

In this chapter an 83-lake cladoceran training set has been created from a subset of the UKAWDDS and the properties of those data analysed. This analysis has two aims: firstly to determine the factors involved in controlling the distributions of Cladocera in acidic fresh waters, and secondly, to determine whether the Cladocera respond to a different set of environmental factors than those important in explaining the distributions of diatom taxa in this particular training set.

From the results presented above, it is clear that the response of the cladoceran taxa to the measured physicochemical is very complex and is related to a range of physical and hydrochemical factors, as well as to a range of unmeasured parameters.

Roughly half of the variance in the Cladocera data can be explained by the 24 physico-chemical factors and a spatial component. The remaining half of the variance is unexplained, though this value is likely to be an over-estimate because of polynomial distortions in the ordination model and the inadequacies of the matrix decomposition when viewed from an ecological point of view (Økland 1999). The unexplained variance can be explained by factors not considered formally in this analysis, such as relative predation pressures from invertebrate or fish sources and habitat structure in the littoral zone (esp. the structure and composition of aquatic macrophyte stands). However, the relative magnitude of and the interrelationships between measured factors and these unmeasured but important factors are largely unknown.

The analysis presented above indicates the important role of hydrochemistry in describing the main patterns in the Cladocera data. The 15 hydrochemical variables uniquely explain almost 20% of the total variance in the data, with a further 6.3% of the total variance being explained by hydrochemistry conditional upon either physical or spatial factors. Calcium (Ca), magnesium (Mg) and TOC explain statistically significant, unique amounts of variance in the cladoceran data. Ca and TOC are correlated with RDA axis 1 whilst Mg is more important on RDA axis three, when only those variables which explain statistically significant amounts of variation are included in the analysis (Table 30 and Figure 43). Ca concentration in acidic surface waters is important because of the role it plays in regulating sodium fluxes into and out of the body, the common cause of mortality in cladocerans in acidic waters (Havas, Hutchinson, and Likens 1984; Zimmer 1987). Ca was also an important variable in explaining the distributions of Cladocera in an alpine training set (Lotter and Birks 1997).

The TOC content of surface waters has been shown to be important in explaining variation is cladoceran abundances (Rautio 2001; Schartau *et al.* 2001). Rautio (2001) has shown that for sub-arctic lakes in Fennoscandia high cladoceran abundances are found in lakes with high humic content despite the low pH of many of these lakes. TOC/DOC acts as filter for harmful short wave ultraviolet radiation, and although this is likely to be more important in lakes in higher latitudes, differences in cladoceran species composition between high and low TOC lakes are seen in the cladoceran training set presented here. TOC is also known to chelate certain metals (such as aluminium) and as a result, acidification, and the concomitant increased mobility of these metals, is likely to have less of an effect in lakes with high TOC than clear-water systems (Nilssen and Sandøy 1986). In a recent study by Schartau *et al.* (2001) in the Sudbury region of Ontario, Canada, DOC in combination with fish species richness and lake area best explained microcrustacean richness in a multiple regression study of lakes on along an acidification gradient.

Physical variables account uniquely for 12.1% of the variance in the Cladocera data with a further 7.2% of the variance explained by physical variables conditional on hydrochemistry and/or spatial factors. Of the twenty-four physico-chemical variables included in the analysis maximum lake depth explained the greatest amount of variance in the Cladocera data (5.2%), a finding that has been demonstrated in a number of regions (e.g. Whiteside

1970; Korhola 1999; Korhola, Olander, and Blom 2000). Lotter *et al.* (Lotter *et al.* 1997) also demonstrate the importance of water depth for benthic (i.e. littoral) Cladocera.

Maximum altitude and net relief in the catchment are the other two physical variables that explain statistically significant unique amounts of the variance in the training set. It is not clear from the results whether this is an effect related to differences in species composition between corrie lakes and other upland systems, or the result of differences between sites due to temperature being expressed in these two variables. The effects of temperature on Cladocera have long been known through pioneering work by Harmsworth (1968), building on earlier work of Goulden (1964), and Whiteside (1970). More recently, attempts have been made to create Cladocera-temperature models to reconstruct temperature from changes in the cladoceran assemblage through time (Lotter *et al.* 1997; Korhola *et al.* 2000). Korhola (1999) demonstrates distinct Cladoceran community compositional changes in 53 sub-arctic lakes along a climate gradient in northern Finland, and Hofmann (2000; 2001) has identified climate induced changes in cladoceran species composition from late glacial/early Holocene sediments from a number of sites, where rapid fluctuations in climate were shown in dramatic changes in the chydorid communities of those lakes.

Temperature may not necessarily directly account for all the variation in the Cladocera data explained by altitude and net relief. Habitat availability, macrophyte structure and predator-prey relationships may also be related to the variance explained by altitude. These factors are likely to be correlated, to some degree, with temperature, but the exact nature of the Cladocera response to altitude is complex and difficult to interpret in this training set without additional data on air or surface water temperatures or more detailed information on habitat structure and predator-prey relationships in the training set lakes.

The results of the RDA of the Cladocera data show concordance with the results of other studies of a similar nature to this one (Korhola 1999). However, the lack of relationship between cladoceran community responses and pH is surprising given the well documented effects of acidity on cladoceran mortality (Krause-Dellin and Steinberg 1984; Arzet *et al.* 1986; Zimmer 1987; Nilssen and Sandøy 1990; Paterson 1994). It is becoming increasingly apparent that in similar fresh water systems to those included in this study, pH has little direct affect on cladoceran community composition (Hofmann 1986; Szeroczynska 1998). The results of this study have shown no relationship with pH in the composition of

Cladocera in the training set after the effects of other, more important variables have been accounted for. The species response curves modelled for pH showed that pH had little effect on the presence or absence of species for the majority of the taxa in the training set. Exceptions to this are *A. guttata* var. *tuberculata*, *A. intermedia*, *M. dispar* and *S. crystallina*. In the case of *A. guttata* var. *tuberculata* and *A. intermedia*, pH relationship is significant ($p < 0.05$) for the unimodal model, as are the responses to alkalinity and calcium, and these two variables lead to a reduction in model deviance that is greater than the reduction in deviance of the pH model. In the case of *M. dispar*, the pH response is significant only for the linear (monotonic) model, and for *S. crystallina*, the pH model is not significant at the 95% level ($p > 0.05$).

Whilst the impact of increasing acidity (such as alkalinity and acid neutralising capacity (ANC) reductions) on cladoceran community composition is not in doubt, a *direct* effect of pH and consequently, therefore, Cladocera-pH inference models and pH reconstructions from cladoceran remains, is questionable. Unlike the diatom-pH response, the pH response in the Cladocera is likely to be seen via *indirect* effects, such as acidity related changes in predator-prey relationships (through loss of or reductions in fish populations in acidified lakes), changes in aquatic macrophyte composition and structure, increased levels of toxic metals (e.g. aluminium), changes in food sources as a result of phytoplankton responses to acidification, as well as the ion imbalances in Cladocera that take place at low pH. The role of calcium and TOC in mitigating the effects of changes in pH is also an important contributory factor controlling the final composition of a Cladocera community at a specific pH. Furthermore, the original biological population supported by a lake prior to the onset of acidification will, in part, determine the changes seen in the Cladocera community through time as that lake acidifies. Lakes in which fish were not present are likely to show different changes in the Cladocera community than those where fish were present prior to acidification, because the relative importance of invertebrate predation in the two lakes will initially be different and will change as acidification progresses.

The analysis of the species response curves for a range of physico-chemical factors indicates that both hydrochemical and physical factors play an important role in explaining the presence or absence of individual taxa in the training set in the same way that a range of physico-chemical factors are important in explaining the broad patterns in the species composition across the whole data set. A range of unimodal, sigmoidal rising or falling,

and linear response curves to a variety of the measured factors can be seen in the results presented above. This suggests that the training set covers long gradients for some variables whilst short gradients are represented for other factors. This appears to be dependent upon particular taxa, with a few taxa showing a unimodal response to pH (see above), whilst the majority have either a non-significant linear response curve or show no response at all to pH. Maximum lake depth is clearly important for a range of chydorid taxa and the response curves for this variable suggest that, for the Chydoridae at least, long lake depth gradients are found in the training set. The planktonic Bosminidae, however, show little response to lake depth in the training set, suggesting that the very shallow lakes are not represented in the training set truncating the response in these taxa.

Cluster analysis of the cladoceran species data has shown the presence of three main community types in the training set: lakes where *B. coregoni* is dominant and chydorids are present at low abundances, lakes where *B. longispina* is dominant and chydorids are present at low abundances, and lake where chydorid cladocerans dominate the assemblage and the planktonic *Bosmina* species are absent or present in low proportions.

Linear discriminant analysis (LDA) has shown that the main differences in the three cladoceran community types can be explained by predominantly physical factors. The first linear discriminant function explains over 70% of the between group variance. Catchment area, lake area and the ratio of these two variables are very important in explaining the differences between the planktonic dominated lakes and the chydorid dominated lakes on the first linear discriminant function. Chloride and magnesium are also important variables in explaining this difference as is TOC. The differences between the two groups of lakes dominated by planktonic *Bosmina* taxa are expressed on the second linear discriminant function. Lakes dominated by *B. longispina* tend to be shallower and have lower conductivities but higher chloride concentrations that those lakes where *B. coregoni* is the dominant taxon.

4.7 Conclusions

It is clear from the results of the analysis of the cladoceran data presented in this chapter that lake depth, conductivity and concentrations of Ca and TOC are the main factors that explain variation in the distributions of cladocerans at all levels of scale in the training set.

At the individual species-level a mixture of hydrochemical and physical factors are important in explaining the responses of particular taxa, a feature which is carried through to the broad compositional level as assessed by RDA. At this broader level, hydrochemical factors explain roughly twice the amount of variance in the training set than physical factors, and maximum lake depth, altitude and relief being the most physical variables and TOC, Ca and Mg being the most important hydrochemical factors in explaining the main patterns in the Cladocera species distributions. At the community type-level, lake size, depth, Ca, TOC and conductivity are the factors that best explain the differences between the three groups of lakes identified on the basis of the cladoceran community the contain.

These data and results suggest a complex interaction between physical and hydrochemical factors at all levels in the training set. No one factor is particularly important in explaining the differences in cladoceran species composition. pH does not have a significant influence on cladoceran community composition in the training set after other, more important, variables are accounted for. Approximately half the variance in the cladoceran data is unexplained, suggesting that other factors, such as habitat availability and structure, and predator-prey relationships are equally important as physical or hydrochemical factors in explaining differences in cladoceran community composition in acidic and acid-sensitive surface waters in the UK. This suggests that caution should applied when interpreting community compositional changes in sediment cores, particularly for pH related changes, but also in any palaeolimnological study, because of the range of factors and complexity of their interaction involved in determining the final Cladocera community record in lake sediments.

The primary aim of this chapter was to investigate whether the factors important in explaining cladoceran distributions in acid sensitive fresh waters were similar to those factors important in explaining the distributions of diatom taxa in similar lakes. The results present above clearly show that the diatom and the Cladocera are responding to different sets of environmental factors. As such, the two groups are complementary from the point of view of analogue matching, and a procedure that uses both diatoms and Cladocera in the calculation of dissimilarity between samples is likely to discriminate between similar and non-similar lakes more faithfully than diatoms alone.

Chapter 5: Applied analogue matching

5.1 Introduction

The technique of analogue matching has been used almost solely for the process of environmental reconstruction over the Quaternary time scale. Analogue matching and the identification of modern analogues is a long established palaeoecological technique for environmental reconstruction from fossil pollen spectra (e.g. Wright 1967; Guiot *et al.* 1989; Fauquette, Guiot, and Suc 1998; Peyron *et al.* 1998). The process of analogue matching involves numerically comparing the degree of (dis)similarity between a sub-fossil assemblage and many potentially analogous modern assemblages. The modern assemblages form a modern training set, containing samples from a wide range of environmental conditions. If the sub-fossil assemblage can be "matched" with one or more of the modern assemblages then it can be inferred that the environmental conditions that occurred at the time the sub-fossil assemblage was created are similar to those operating where the modern assemblage was collected. The matched modern samples are known as modern analogues. Until recently, the type of assemblage compared has predominantly been restricted to fossil pollen spectra, though potentially any assemblage could be used given the constraint that the assemblage in question is present in the modern and the fossil environment.

However, the use of the technique can be applied to the more recent past in order to investigate the problems of recent anthropogenic environmental disturbance (Flower *et al.* 1997). Flower *et al.* (1997) examine the potential for using the analogue matching approach to set restoration targets in acidified surface waters in northwest Europe. By comparing the sub-fossil remains of diatoms to the modern diatom assemblages from a modern training set of acid and acid-sensitive lakes from across northwest Europe, Flower *et al.* (1997) were able to identify modern analogue lakes for the two acidified surface waters they

examined. Their contention was that these modern analogues could be used to set baseline reference conditions for the impacted lakes, and potentially be used as targets to evaluate the future restoration of two acidified lochs in the light of the reductions in the emission of acid forming compounds across Europe.

Whilst demonstrating the value of the approach the Flower *et al.* (1997) study illustrated a number of problems. The closest modern analogues identified from the modern training set were found to not be the most pristine sites in the modern training set. These sites are 'minimally impacted' lakes, certainly with regard to their diatom floras. A further issue with the two closest modern analogues was that the lake-water calcium concentrations for the two lakes was much higher than the modern day calcium values measured at the two studied lochs. Leaching of base cations is enhanced by acidification leading to increased base cation concentrations in runoff and increased H^+ concentrations in soils (Henriksen, Kamari, and Wilander 1992). As acid deposition declines the amount of deposited H^+ ions stored in the soil and, therefore, exchanged for base cations is reduced, which leads to a reduction in base cation concentrations in runoff. Therefore, because Ca concentrations in lake waters will decline as acid deposition is reduced the lake water Ca concentration for the reference condition will be lower than the Ca concentration recorded at the present day. The mismatch between the likely Ca concentrations of those suggested by the two closest analogues is particularly significant given the importance of Ca in the distributions of many fish and invertebrate species (Muniz and Walløe 1990). The insensitivity of the diatoms to changes in calcium concentrations, therefore, is a serious flaw in the diatom-based methodology.

5.2 Analysis of the diatom and cladoceran analogue matching data set

In the previous two chapters the distributions of the diatoms and the Cladocera in the various training sets have been analysed using a range of multivariate numerical techniques. Using constrained ordination techniques, the relationships between the species and the physico-chemical variables were investigated. The results of these analyses has shown that the two species groups are responding to or reflecting different aspects of the physico-chemical environment in the training set.

This is an important point from the point of view of analogue matching and the development of the new approach that will use both diatoms and Cladocera in the matching process. Having shown that the diatoms and the Cladocera respond to different aspects of the physico-chemical environment it is likely that the matching process will incorporate these differences, and thus provide more robust matches than the original diatom approach (Flower *et al.* 1997).

Prior to commencing any further matching it is important to demonstrate that when the diatom and the cladoceran data sets are combined into a single multiproxy analogue matching training set that the properties of the individual data sets are preserved.

The method used to join the diatom and Cladocera data sets together is described in the methods chapter of the thesis (see section 2.4.2 above). DCCA OF the joint data set showed that the species data exhibit short gradients of less than 2 Hill's standard deviation units. The short gradient length indicated by DCCA suggests a linear response of the species to the underlying environmental gradients and that the use of redundancy analysis (RDA) is appropriate in this case.

Table 38: Summary results of the RDA of the joint diatom and cladoceran data sets

Axes	1	2	3	4	Total Variance
Eigenvalue (λ)	0.081	0.049	0.029	0.020	1.000
Species – Environment correlation	0.718	0.729	0.686	0.523	
Cumulative % variance of species data	8.1	13.1	15.8	17.8	
Cumulative % variance of species environment correlation	42.2	67.6	82.0	92.3	
∑ all unconstrained λ					1.000
∑ all canonical λ					0.193

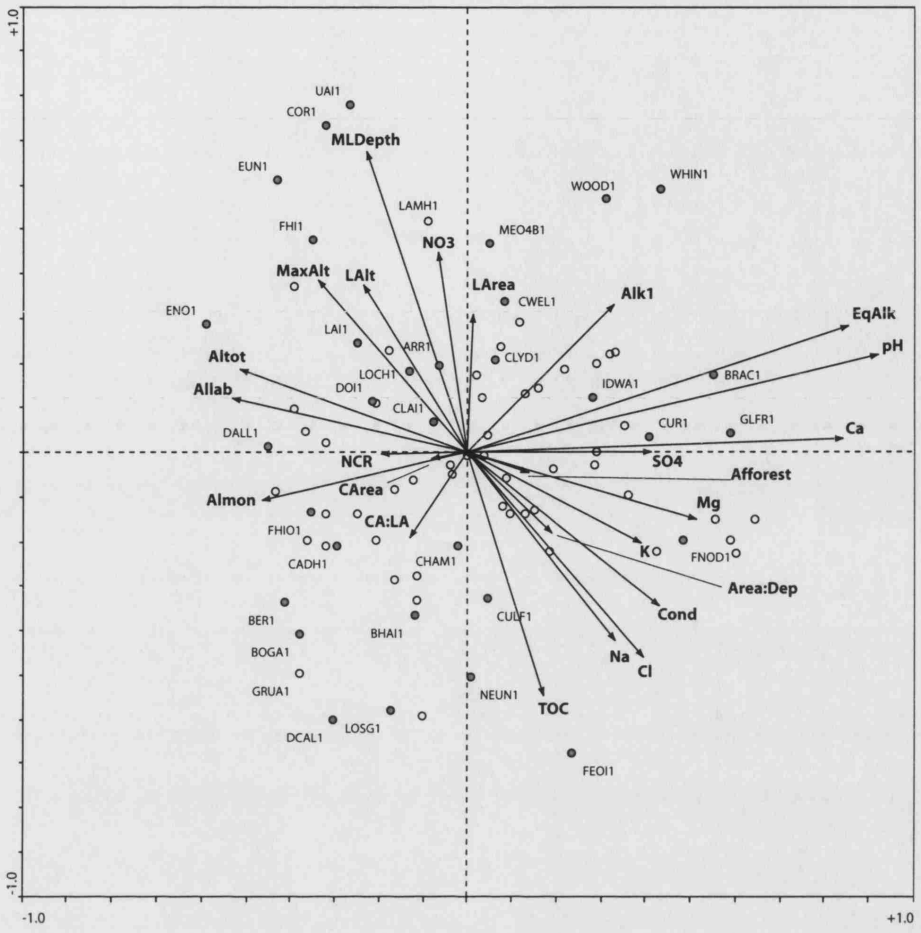


Figure 70: RDA biplot showing the relationship between the 83 lakes in the joint diatom and cladoceran data set and the measured physico-chemical variables

Table 38 shows the summary results of the RDA using all 24 physico-chemical variables. Almost 27% of the variance in the species data is explained by these 24 environmental variables on the first 4 axes of the RDA, and that 57% of the relationship between the species and the measured environmental variables can be explained by the 4 axes of the RDA, the 4 hypothetical gradients.

Figure 70 shows the RDA biplot of samples and environmental variables for the joint data set. It is clear from this biplot that a strong acidity gradient is present in the joint data set. Long biplot arrows are indicated for pH, equivalent alkalinity and calcium, which show a strong positive correlation between these variables and the main species-environment pattern in the data set. The biplot arrows for the three aluminium variables (labile, monomeric and total) indicate strong negative correlations with the first RDA axis. This is in contrast with the results of the 163-lake diatom data set, which show that aluminium is

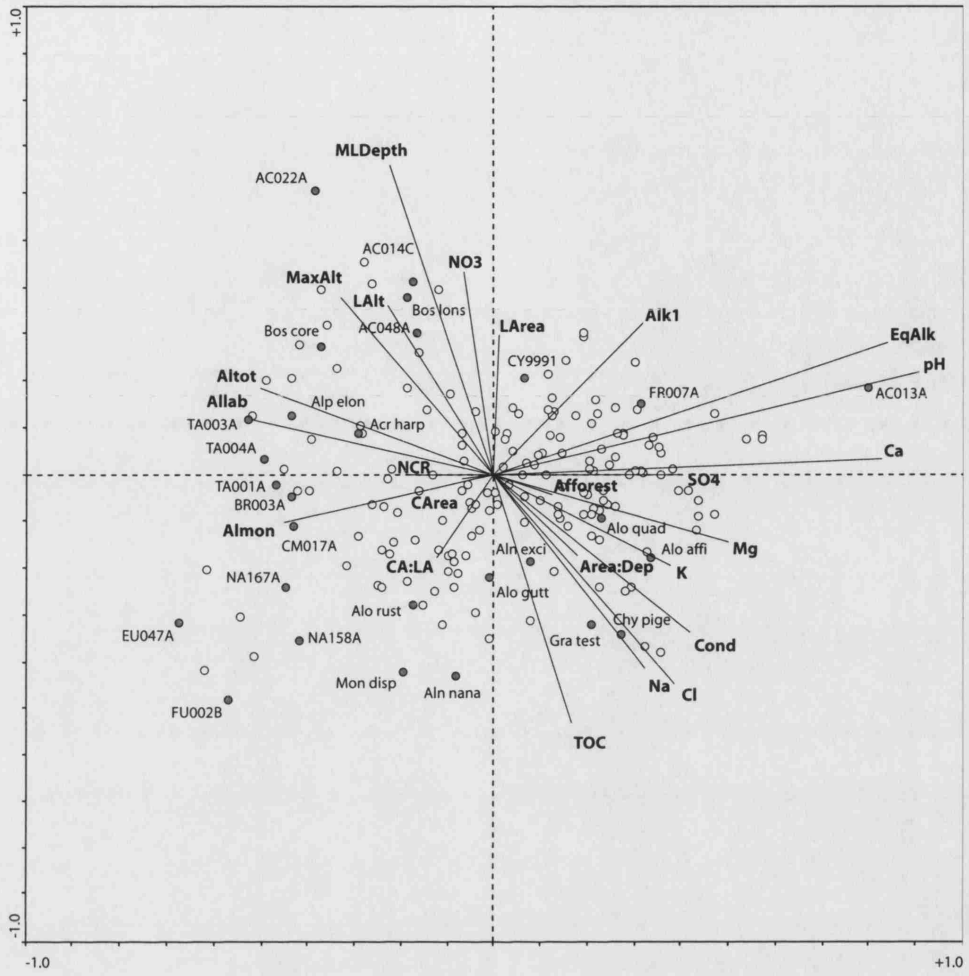


Figure 71: RDA biplot illustrating the species-environment relationships in the joint diatom and Cladocera training set.

strongly correlated with both the first and second ordination axes (See section 3.2.3 above, and Figure 19). The CCA of the 83-lake diatom training set does not show the same correlation in aluminium with the second axis of the CCA. In this respect the joint data set and the 83-lake diatom training set are quite similar.

Figure 71 shows the RDA biplot of the species environment relationships in the joint diatom and Cladocera training set. The first axis illustrates the strong diatom-pH relationship, with acid sensitive taxa, such as *A. minutissima* (AC013A), associated with high pH lakes on the right of the plot and with acid tolerant taxa, such as *T. quadriseptata* and *T. binalis*, associated with low pH lakes on the left. On the second RDA axis there is a strong contrast between samples with high proportions of chydorid *Cladocera* associated with the shallower, high conductivity and high TOC lakes at the bottom of the biplot, and samples

dominated by planktonic *Bosmina* species in the deeper lakes towards the top of the diagram.

The variance inflation factors for the physico-chemical variables reported by CANOCO for both of the training sets indicate that a large degree of multi-colinearity in the environmental data. In such datasets the regression or canonical coefficients are unreliable and the inter-set correlations of the environmental variables with the ordination axes give a better indication as to which variables are most associated with the constrained ordination axes. The inter-set correlations of the first ordination axis in both training sets confirm that this axis is a strong acidity gradient (Table 39). In both training sets, pH is the most strongly correlated variable with the first ordination axis (inter-set correlation = 0.8397 [diatom] and 0.8493 [joint]), with equivalent alkalinity (inter-set correlation = 0.7803 [diatom] and 0.7857 [joint]) and calcium (inter-set correlation = 0.7714 [diatom] and 0.7753 [joint]) also strongly correlated with this axis.

Whilst the first constrained ordination axis is very similar in both training sets, the second and subsequent axes in the data sets are quite different. TOC is the variable that is most strongly correlated with the second axis of the CCA of the 83-lake diatom data set (inter-set correlation = -0.5705), whilst maximum depth is the variable most closely related to RDA axis 2 in the joint data set (inter-set correlation = 0.5998). TOC is also negatively correlated with this RDA axis but the correlation (-0.4844) is somewhat lower than in the CCA of the diatom data. This indicates greater importance of maximum lake depth in describing an important pattern in the diatom and cladoceran data than in the diatom data set where maximum lake depth is only weakly correlated with the second constrained ordination axis (0.3682). In both data sets chloride and nitrate concentrations are also important variables on the second axis of variation.

The third axis of the CCA of the diatom training set is closely correlated with maximum lake depth (-0.5623) and lake area (-0.4707) whilst net catchment relief is the most important variable on the third axis of the RDA of the joint data set. Maximum lake depth is uncorrelated (-0.060) with the third RDA axis in the joint data set further indicating that this variable is more important in explaining the main patterns in the species data from the joint data set than in the diatom data alone. The correlation between net catchment relief and the third RDA axis suggests that the diatoms and Cladocera in high mountain corrie

lakes (where the net relief in the catchment is likely to be high) different from those sites with less relief in their catchment. This might indicate a light effect on the species or differences in habitat availability across the range of sites rather than a direct altitude effect in this case as lake altitude in uncorrelated with this axis (0.0045). Whatever the reasons for this pattern in the diatom and cladoceran data it is a feature that is emphasised to a larger extent in the joint data set than in the diatom data set, where net catchment relief is more related to the 4th axis of variation than the 3rd.

The preceding analysis has demonstrated that when the diatom and cladocera data sets are combined into a single training set, the properties of the two individual data sets remain expressed in the combined training set. The primary axis of variation in the combined training set is still an acidity-related gradient; the variables that are closely correlated with this axis are only loosely correlated with the secondary axis of variation. In the diatom training set the acidity-related variables were also closely correlated with the secondary axis of variation. A property of this combined data set is that the secondary axis of variation (RDA axis 2) is a gradient contrasting large, deeper lakes at higher elevations with high conductivity and TOC lakes. This is a complex gradient but illustrates the increased importance of a wider range of variables in the combined training set. This will be important for the analogue matching procedure, where the distances between lakes in multivariate space are related to a much wider range of factors than the pH – conductivity response seen in the diatom-only training set.

Table 39: Inter-set correlations of environmental variables with ordination axes for the 83-lake diatom training set and the joint diatom and Cladocera training set.

Variable	83-lake Diatom training set(CCA)				83-lake Joint training set (RDA)			
	Axis 1	Axis 2	Axis 3	Axis 4	Axis 1	Axis 2	Axis 3	Axis 4
Alkalinity	0.2728	0.2712	0.2364	0.2874	0.2990	0.2909	0.1216	-0.1315
Labile aluminium	-0.5082	0.1741	0.0657	0.0938	-0.4843	0.1071	0.1195	-0.2871
Monomeric Aluminium	-0.4159	-0.0082	0.0131	-0.0918	-0.4217	-0.0939	-0.0349	-0.0844
Total aluminium	-0.4621	0.1797	-0.653	0.2687	-0.4710	0.1652	0.3122	-0.1497
Calcium	0.7714	-0.0225	0.565	0.2888	0.7753	0.0280	0.2747	-0.0155
Chloride	0.4042	-0.4336	-0.0708	-0.2657	0.3629	-0.4086	-0.2268	0.1912
Conductivity	0.4249	-0.3479	-0.0266	0.0178	0.3938	-0.3045	0.0579	0.0240
Equivalent alkalinity	0.7803	0.2305	0.0248	-0.0543	0.7857	0.2545	-0.0720	0.1289
Potassium	0.3687	-0.1687	-0.0213	-0.0581	0.3531	-0.1798	-0.0672	0.0284
Magnesium	0.4777	-0.1832	0.0422	-0.986	0.4691	-0.1316	-0.1623	0.0336
Sodium	0.3422	-0.3938	-0.0912	-0.1939	0.3030	-0.3738	-0.1710	0.1603
Nitrate	-0.1114	0.4063	0.0698	0.3274	-0.0597	0.3929	0.2463	-0.2162
pH	0.8397	0.2129	0.0278	-0.1094	0.8493	0.1953	-0.0853	0.1128
Sulphate	0.3533	-0.736	0.860	0.3625	0.3701	-0.0011	0.2255	-0.1691
TOC	0.1795	-0.5705	0.0605	0.2476	0.1558	-0.4844	0.3787	0.0374
Afforestation	0.1508	-0.0368	-0.2458	-0.1217	0.1168	-0.0394	-0.0864	0.3995
Area depth ratio	0.2094	0.1547	-0.1950	0.1248	0.1677	-0.1559	0.1628	0.3164
Catchment lake ratio	-0.1370	-0.0780	0.0707	-0.2943	-0.1154	-0.1633	-0.2551	0.0865
Catchment area	-0.0326	-0.0261	-0.2016	-0.1851	-0.0660	-0.0110	-0.1979	0.3151
Altitude	-0.1971	0.2624	0.0655	0.0896	-0.2142	0.3266	-0.0045	-0.2487
Lake area	0.0816	0.1555	-0.4707	0.0878	0.0107	0.2696	0.0591	0.4905
Maximum altitude in catchment	-0.2854	0.3890	0.0548	-0.2417	-0.3102	0.3417	-0.3257	0.0126
Maximum depth	-0.1285	0.3682	-0.5623	0.0374	-0.2110	0.5998	-0.0600	0.4667
Net Catchment relief	-0.1269	0.0597	-0.1033	-0.4837	-0.1690	-0.0029	-0.4843	0.2206

5.3 Ecological resemblance

The basis of analogue matching is ecological resemblance. The similarity of any two samples is measured by (dis)similarity coefficients or distance metrics, which are quantitative descriptors of the ecological resemblance between two samples. There are many ways of formulating measures of resemblance, each way emphasising different aspects of the assemblages being compared. This section of the thesis introduces the concept of ecological resemblance and discusses suitable dissimilarity coefficients for use in analogue matching using diatom and cladoceran remains in fresh waters.

There are three types of ecological "resemblance"; similarity, distance, and dependence. Measures of resemblance are used routinely in many ecological analysis techniques including ordination and classification (cluster analysis).

We can look at relationships between either the objects (sites) or the descriptors (e.g. species or environmental variables). These two types of resemblance are known as Q-mode and R-mode respectively. Cattell (1966) takes this one step further, and observes the data box of objects x descriptors x time instances (Figure 72), and defines 6 modes of analysis.

Table 40: The six modes of analysis as defined by Cattell (1966) as shown in the three-dimensional data box (Figure 72).

Mode	Description
O	Among time instances, based on all descriptors for a single object
P	Among descriptors, based on all time instances for a single object
Q	Among objects, based on all descriptors for a single time instance
R	Among descriptors, based on all objects for a single time instance
S	Among objects, based on all time instances for a single descriptor
T	Among time instances, based on all objects for a single descriptor

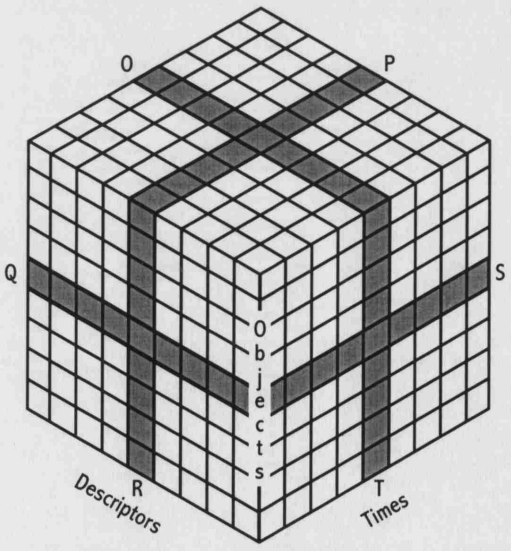


Figure 72: (Right) The three-dimensional data box (objects × descriptors × times). Adapted from Cattell (1966).

O-mode analyses tend to be conducted using Q-mode coefficients. Similarly, P-mode analyses use R-mode coefficients except where P is a time series, in which case special R-mode coefficients are used. S- and T-mode studies are mostly autecological involving only a single species. S-mode analyses are conducted using P-mode coefficients. T-mode studies use statistical tests of hypotheses for related samples, and where two time instances are to be compared, they can be considered as two descriptors and analysed as for S-mode. Environmental Impact studies are one major form of T-mode analysis, such as BACI designs (Before, After, Control, Impact) (Legendre and Legendre 1998).

The usual way of representing resemblance is to condense the information contained in an ecological data matrix into a square matrix of association among objects or descriptors. In most cases association matrices are symmetric. For example, association matrices are produced as the output of clustering and ordination techniques. These results do not necessarily reflect all of the information contained in the original data matrix. It is important therefore, to choose an appropriate measure of association. The chosen measure of association must take into account the nature of the study and the research questions to be tackled or the hypothesis to be tested. Furthermore, the different measures of association matrix might call for a measure of association with certain mathematical properties.

In Q-mode analysis we can make the distinction between similarity coefficients and distance (or dissimilarity) coefficients. Similarity coefficients reach a maximum value when two objects are perfectly identical and a minimum when the two objects are completely different. Conversely, distance coefficients follow the inverse of this rule.

5.3.1 The Double Zero Problem

The double zero problem occurs in ecology as the result of the special nature of the species descriptors. Species exhibit unimodal distributions along environmental gradients, occurring most abundantly at their optimal environmental conditions. Species become rarer, and will eventually disappear, as environmental conditions diverge from the optima. If a species X is present in object A and object B then this is correctly identified as a similarity between the two objects. The inverse of this statement does not hold however, as species X might be absent from both objects A and B but for different autecological

reasons. For example, object A might be found at the upper end of the environmental gradient, whilst object B might occupy a position at the lower end of the gradient. If species X has its optima around the centre of the environmental gradient then it may be absent from both A and B but clearly A and B differ in their environments. So ecologists try not to draw any ecological conclusions from the absence of a species in two objects. Therefore, our measure of resemblance should ignore double absences (known as double zeroes) in their mathematical formulation. Measures of resemblance which ignore double zeroes are *asymmetrical* because of the differences between the way zeroes are treated and the way other numbers are handled in the calculation.

5.3.2 Q-mode similarity coefficients

These resemblance measures are by far the largest group of coefficients found in the ecological literature. Similarity coefficients are not "metric" because it is possible for there to be two objects A and B that are more similar than the sum of their similarities with another, more distant object C. As a result they cannot be used in ordination techniques, but they have been routinely used in cluster analysis and classification problems.

5.3.3 Q-mode distance measures

Distances, like similarities, are used to represent the resemblance between two objects. There are three types of distance coefficients: metric, semi-metrics and non-metrics. Metrics follow four rules: firstly if object A = object B then the distance between them must equal zero as well. Secondly, if A is not equal to B then the value taken by the distance measure must be positive. Thirdly, metrics are symmetric in that the distance between A and B must be the same as the distance between B and A. Fourthly, they must follow triangle inequality, such that the distance between A and B plus the distance between B and C must be greater than or equal to the distance between A and C (Legendre and Legendre 1998).

Semi-metrics, also known as pseudo-metrics, do not follow the triangle inequality property and therefore cannot be used to display points in metric or Euclidean space because the distances from A to B and from B to C may be smaller than the distance between A and C. This is the same reason why Q-mode similarity coefficients cannot be used in ordination

techniques. Non-metrics violate the second rule in that they take negative values (Legendre and Legendre 1998).

5.3.4 The dissimilarity matrix

An association matrix containing the degree of resemblance between each object and every other object (Q-mode) is produced by calculating the distance or similarity between each object in the ecological data matrix and each other object. This association matrix is known as the dissimilarity (distance) or similarity matrix. For n objects there are $1/2n(n-1)$ resemblance measures to be calculated. A dissimilarity matrix may be presented as a full square matrix, where the number in the cell in matrix M representing the resemblance between objects x and y (M_{xy}) takes the same value and the number in the cell M_{yx}. For a dissimilarity (distance) matrix the values on the diagonals will be zeroes. Consequently, the square dissimilarity matrix contains a redundant amount of information. Generally resemblance matrices are presented as triangular matrices, with either the upper or lower triangle of the square resemblance matrix presented. Diagonal values are also generally omitted (Legendre and Legendre 1998).

5.3.5 Properties of some dissimilarity coefficients

Prentice (1980) and Overpeck *et al.* (1985) investigate the properties of a range of dissimilarity/distance coefficients and their applicability for use in multidimensional scaling, other ordination techniques and the interpretation of fossil pollen spectra. These studies identified three types of dissimilarity coefficient; simple or unweighted dissimilarity coefficients, equal weight coefficients, and signal-to-noise coefficients. The three types of dissimilarity measures differ in the importance attached to the less common taxa in the objects, as well as in the weight given to individual, large differences in species abundances as against accumulated small differences in abundance between objects.

Simple dissimilarity coefficients do not use any weighting of the species abundances. As a result these measures are known to be overly dominated by the most common taxa, with differences between any two objects in the abundances of the rare taxa carry little weight.

Equal weight dissimilarities counter the dominance of the most abundant taxa by proportionally up-weighting the less common taxa.

Signal to noise measures extract the signal or pattern in the species data from the noise contained within the data. The noise is the random variation in the abundances of the taxa in the data matrix. It is assumed that these random effects act equally over all objects and at all times, and that they lead to a distribution of taxon abundances that is proportional to a multinomial distribution. All of the signal-to-noise coefficients in Table 41 imply a flexible weighting of the less common taxa such that those taxa that exhibit wide ranges could be treated as common in some analyses and rare in others (Overpeck *et al.* 1985).

5.3.6 Choice of dissimilarity coefficient

Manhattan and Euclidean distance coefficients are widely used in fields such as palynology, where the use of analogue matching approaches to palaeoenvironmental reconstruction are far more advanced. These two dissimilarity coefficients are attractive because of their simplicity. They attach importance to each taxon based on its abundance, but because of this they are known to be overly influenced by the most abundant taxa in the data matrix.

Table 41: Table showing the properties of some common dissimilarity coefficients (sources; Overpeck et al. (1985) & Legendre and Legendre (1998)

	Name	Formulation	Notes
Simple coefficients	Manhattan Metric	$d_{ij} = \Sigma_k \left\| p_{ij} - p_{jk} \right\|$	This coefficient measures the absolute difference between the species abundances for each species found in two objects
	Euclidean distance	$d_{ij} = \Sigma_k \left(p_{ij} - p_{jk} \right)$	This coefficient measures the sum of the squared differences between the species abundances found in any two objects. Implicit in ordinary PCA.
	Cosine theta similarity	$d_{ij} = \Sigma_k \left(\frac{p_{ij}}{\left(\Sigma_k p_{ik}^2 \right)^{1/2}} - \frac{p_{jk}}{\left(\Sigma_k p_{jk}^2 \right)^{1/2}} \right)$	This coefficient is the Euclidean distance between any two objects after normalizing the species vectors.
Equal weight coefficients	Canberra Metric	$d_{ij} = \Sigma_k \frac{\left\| p_{ij} - p_{jk} \right\|}{p_{ik} + p_{jk}}$	The Canberra metric is the sum of a series of fractions, one per taxa, of the abundances in any two objects. As such it cannot be dominated by the major taxa. However, it is overly dominated by species presences versus absences and double zero problems.
	Standardized Euclidean Distance*	$d_{ij} = \Sigma_k \left(\frac{p_{ik} - p_{jk}}{s_k} \right)$	This is the Euclidean distance computed on standardised (zero mean, unit standard deviation) species data. Used implicitly in standardized PCA.
	Gower's coefficient*	$d_{ij} = \left(2\Sigma_k \frac{\left\| p_{ik} - p_{jk} \right\|}{R_k} \right)^{0.5}$	Gower's coefficient is a standardized Manhattan metric. The standardization implied is standardising by the range of proportions for taxon *k*.
	**: Note that the standardized Euclidean Distance and Gower's coefficient actually up-weight those taxa that vary within narrow frequency limits rather than rare taxa per se. These two methods tend to up-weight rare taxa as these species tend to have restricted ranges.*		
Signal to noise coefficients	Chord distance	$d_{ij} = \Sigma_k \left(p_{ik}^{0.5} - p_{jk}^{0.5} \right)$	This coefficient represent the length of a chord drawn between the two vectors in a hyper-dimensional space (described by n square root transformed species proportions). It can be seen that the chord distance equates to a Euclidean distance of square root transformed data.
	Information Statistic	$d_{ij} = \Sigma_k \left(p_{ik} \ln \frac{2p_{ik}}{p_{ik} + p_{jk}} + p_{jk} \ln \frac{2p_{jk}}{p_{ik} + p_{jk}} \right)$	The information statistic is a function of the likelihood ratio for distinguishing two multinomial distributions with equal taxon sums.
	Angular separation	$d_{ij} = \cos^{-1} \Sigma_k \left(p_{ik} p_{jk} \right)^{0.5}$	Angular separation describes the angle between two vectors in a polar co-ordinate system defined by the transformed frequency of the taxa.
	χ^2 coefficient	$d_{ij} = \left[\Sigma_k \frac{\left(p_{ik} - p_{jk} \right)^2}{p_{ik} + p_{jk}} \right]^{0.5}$	The squared χ^2 coefficient is equivalent to Mahalanobis' D^2 applied to multinomial distributions with equal taxon sums.
	Where, d_{ij}= dissimilarity between object *i* and *j*; p_{ik} = the proportion of taxon k in object *i*; R_k = range of proportions for taxon *k*; sk = standard deviation of proportions for taxon k.		

Of the equal weight measures described in Table 41, standardized Euclidean distance is an attractive coefficient for use in describing ecological resemblance as it is the dissimilarity measure preserved in standardised PCA (PCA of a correlation matrix). Equal weight measures are sensitive to changes in the number of rare taxa found in the objects. For example, standardised Euclidean distance leads to a large weighting of the rare taxa and the measure also emphasises slight variations between common taxa over objects so that they have equal importance to the large scale differences between objects. This sensitivity can lead to an emphasis of the noise or random variation at the expense of the signal or pattern in the data.

Signal to noise measures fit into the space between simple dissimilarity coefficients and equal weight coefficients. They avoid the difficulties associated with the equal weight measures, implying a lesser weighting of the rare taxa than these measures, but they weight the rare to a greater degree than the simple dissimilarity coefficients do. Signal to noise measures call for assumptions to be made about the nature and distribution of the taxon abundances being compared, but these assumptions are not overly restrictive, and many of them are indeed implicit in many of the other forms of dissimilarity coefficients.

The signal to noise coefficients described in Table 41 are all formulated in a similar when looked at from first principles. The chord distance and the chi-square distance measures are suitable for analysing absolute taxon abundances as well as the formulations for proportional abundances described above.

Various criteria for choosing a dissimilarity coefficient lead to different choices of coefficients. For simplicity the Manhattan metric or Euclidean distance measure are recommended, though the problems with double zeroes and their insensitivity to differences in rare taxa should be recognised (Legendre and Legendre 1998). Standardised dissimilarity coefficients should be avoided if possible because of the excessive weightings given to the rare taxa (Overpeck *et al.* 1985). Standardized Euclidean distance is still a popular choice for measuring ecological resemblance as it is the dissimilarity measure preserved in standardised PCA (Overpeck *et al.* 1985). Signal to noise measures have been shown to perform well in tests on pollen spectra and have the desirable property of emphasising the main patterns in the data (Overpeck *et al.* 1985). All the measures described in Table 41 obey the upper limit rule, in that they attain an upper limit if, and

only if, two objects are perfectly dissimilar (Overpeck *et al.* 1985). As already stated, these signal to noise measures can be formulated in terms of each other from first principles, so despite not appearing to have very much in common, they can all be expected to perform similarly well on the same species data (Legendre and Legendre 1998). If we combine the desirable properties of simplicity and emphasis of signal over noise, then the chord distance combines these properties into a useful dissimilarity coefficient shown to perform well in tests on pollen spectra (Overpeck *et al.* 1985).

5.3.7 Comparison of distance measures

Overpeck *et al.* (1985) have shown theoretically that signal-to-noise coefficients have desirable properties for determining ecological resemblance amongst samples. They show that the signal-to-noise coefficients also perform better than simple and equal weight coefficients, when comparing modern pollen spectra from North America with fossil samples from cores across the Northern United States of America.

Pollen spectra and the sub-fossil remains of diatoms and Cladocera are two very different entities. Whilst the squared chord distance might be appropriate for comparing pollen spectra it has not been established which of the coefficients is most appropriate for comparing diatom and cladoceran sub-fossil remains. Before attempting to identify modern analogues for acidified systems in the UK, therefore, the ability of the respective coefficients to distinguish between different lakes on the basis of their sub-fossil diatom and cladoceran remains should be investigated. This task involves two stages.

The first stage is concerned with how we determine whether any given sample is similar to any other sample. Classically this has been achieved through the use of cut-off or critical values for the similarity coefficients. These cut-off or critical values describe the level of resemblance between any two samples before they are classified as being "similar" to each other. It might also be desirable to define more than one critical value, and in doing so, introduce the concepts of "close", "very-good" and "good" analogues. So now we have the ability to distinguish between the most similar samples, and those that are less similar but still meet our resemblance criterion.

After establishing critical values for our dissimilarity coefficients, each coefficient needs to be assessed in terms of its ability to discriminate between samples that are known to be similar in, say, their physical and hydrochemical properties and those that are not. For this we might choose to take our modern training set of samples, and identify important gradients or axes of variation in them, and then look to see how our measures of resemblance perform along those gradients. Looking at our main gradients in this way might also indicate where along those gradients that suitable coverage of the environmental space is described by the modern training set and where the coverage might not be so complete.

These critical values for the coefficients can be determined in a number of ways. The simplest way of establishing critical values for a coefficient is to choose arbitrarily these values, based on the known properties of the resemblance measure. In the case of the Squared Chord dissimilarity measure, values of which fall between 0 and 2, it might be decided that suitable close analogue sample should differ from another sample by a factor of 0.1 or 10%, very-good analogues by a factor of 0.2 and good analogues by 0.3. So in these terms, critical values for the squared chord distance would be set at 0.2, 0.4, and 0.6 for close, very-good and good analogues respectively. This method is simple and can be based on expert knowledge, yet its major disadvantage is that it is difficult to translate expert knowledge into a numerical measure of how similar any two samples should be for them to be classed as close analogues. In these situations these critical values become arbitrary and may bear little resemblance to the actual levels of similarity required.

A second way of establishing critical values for the measures of resemblance is to generate them empirically from the observed distribution of resemblance values of every sample in the modern training set with every other sample (Birks pers. commun). One might then examine the observed distribution and determine the required critical values from the values observed. This might be impractical when the modern training set is large, so a more robust way of establishing the empirical critical values would be to calculate the extreme percentiles of the distribution. Figure 73 and Table 42 show the empirical critical values for the various training set discussed in the thesis, derived by taking percentiles of the distribution of resemblance values when each sample in the modern training set is compared with every other sample.

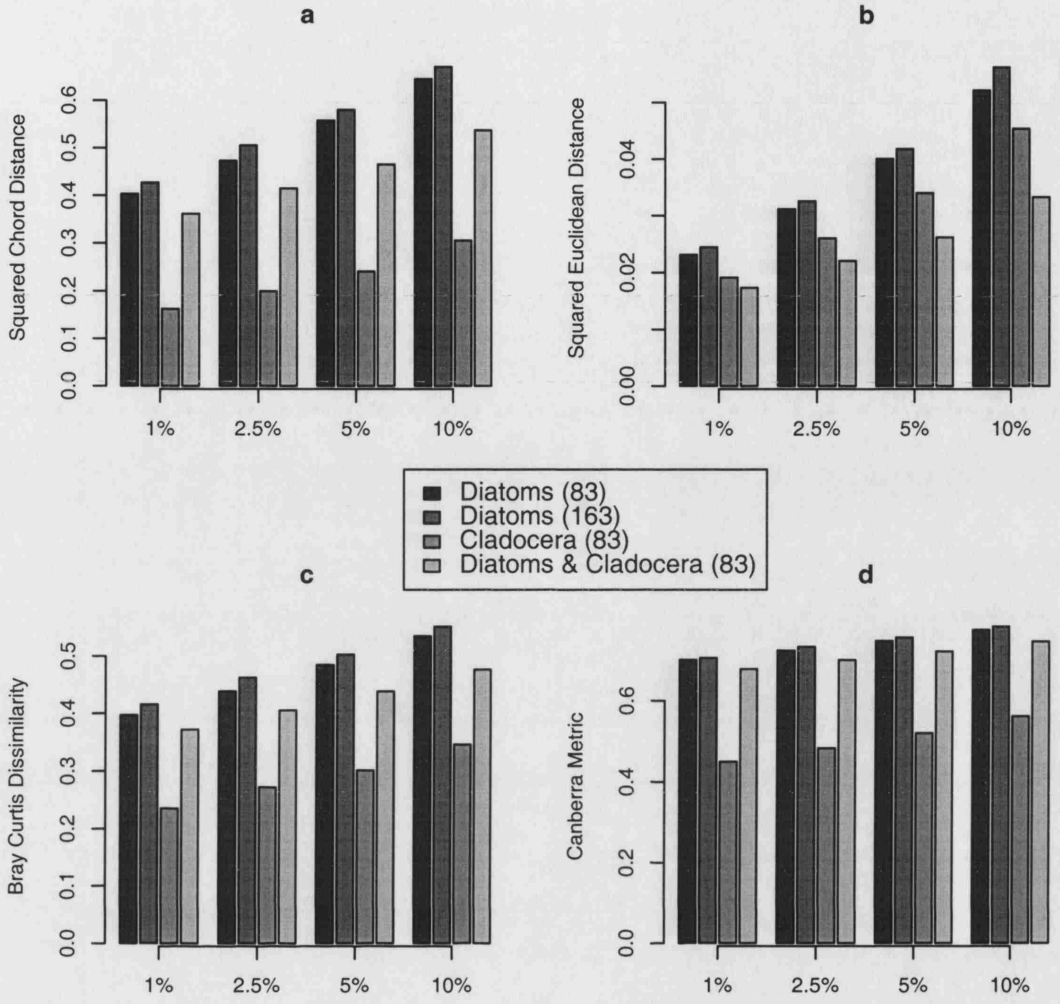


Figure 73: Critical Values for the Squared Chord Distance (a), Squared Euclidean Distance (b), Bray Curtis Percentage Dissimilarity (c) and Canberra Metric (d) resemblance measures determined for the fours analogue matching data sets. The numbers in the brackets in the legend are the numbers of samples in each training set.

A decision has to be made as to which of the extreme percentiles should be used as the critical values for close, very-good and good analogues respectively, but useful percentiles might be the 1st, the 2.5th, the 5th and the 10th. These percentiles indicate the critical value needed for any two samples to be as similar as 1%, 2.5%, 5% and 10% of the samples in the modern training set are to another single sample. The advantage of using empirical critical values is that the data indicate the level of resemblance required. These empirical critical values are heavily influenced by the shape of the observed distribution of the dissimilarity values, however, which may lead to higher or lower than appropriate critical values being established (see below, section 5.3.7.2).

Table 42: Critical Values for the Squared Chord Distance, Squared Euclidean Distance, Bray Curtis Percentage Dissimilarity and Canberra Metric resemblance measures determined for the four analogue matching data sets. The 1st, 2.5th, 5th and 10th percentiles of the observed distribution of resemblance measures are shown.

	1st	2.5th	5th	10th
Squared Chord Distance				
Diatoms (83)	0.4033389	0.4730977	0.5577743	0.6452646
Diatoms (163)	0.4270099	0.5054121	0.5802107	0.6709961
Cladocera (83)	0.1623099	0.1996681	0.2407058	0.3056275
Diatoms & Cladocera (83)	0.3612642	0.4147594	0.4651734	0.5376247
Squared Euclidean Distance				
Diatoms (83)	0.02312377	0.036116154	0.04006321	0.05217105
Diatoms (163)	0.02444326	0.03256566	0.04180849	0.05626964
Cladocera (83)	0.01912875	0.02600662	0.03400979	0.04541328
Diatoms & Cladocera (83)	0.01734406	0.02204396	0.02618948	0.03329049
Bray Curtis Percentage Dissimilarity				
Diatoms (83)	0.3973251	0.4384450	0.4844706	0.5350367
Diatoms (163)	0.4160630	0.4621499	0.5028636	0.5518285
Cladocera (83)	0.2356658	0.2722118	0.3014911	0.3461873
Diatoms & Cladocera (83)	0.3720843	0.4053494	0.4382828	0.4764002
Canberra Metric				
Diatoms (83)	0.7015538	0.7251659	0.7488331	0.7763609
Diatoms (163)	0.7063677	0.7343195	0.7573351	0.7845520
Cladocera (83)	0.4503714	0.4834773	0.5202298	0.5622186
Diatoms & Cladocera (83)	0.6789834	0.7012908	0.7228866	0.7475790

Related to the empirical method of establishing the critical values, is a computer intensive permutation-based method. In its most simple form, one could simply permute at random the values in the objects by species matrix to generate a set of random species data. Take at random a single sample pair and calculate the resemblance of the two samples using a suitable resemblance coefficient. Repeat this process n times, each time selecting a single pair of samples at random and calculating the resemblance between the two samples. For the distribution of the n resemblance values calculate the extreme percentiles as required and use these percentiles as the critical values. A suitable value for n would be $\frac{1}{2}N(N-1)$, where N is the number of samples in the original training set. Using this number of random values results in a random set of resemblance values that is the same size as the lower (or upper) triangle of the resemblance matrix.

An immediate problem with the approach described above is that there are no constraints on the permuted data. The permuted data is just as likely to produce samples that are dominated by a taxon that is of low abundance in the modern training set, thus resulting in random species data that is very unlikely to have resulted in nature if all the available environmental space had been sampled. To counteract this, constraints can be introduced such that only values within a column of the objects by species matrix are permuted and not the whole matrix. Now, any value in a given column can only take a value observed in that column. In this way, those taxa that are naturally rare species remain rare in the permuted data.

Another way of generating the critical values is to use a random number generator to generate random, normally-distributed deviates that have the same range and the same minimum and maximum values taken by the relevant resemblance measure. The extreme percentiles can be determined from this distribution and they can either be used directly as the critical values, or to evaluate the suitability of the empirical critical values established from the observed distribution or the permuted distribution of resemblance values. This method can only be performed when the resemblance measure takes a maximum value when the two samples are perfectly dissimilar (or similar), such as the squared chord distance. The Euclidean distance on the other hand does not take a maximum value and so randomly generated critical values cannot be determined.

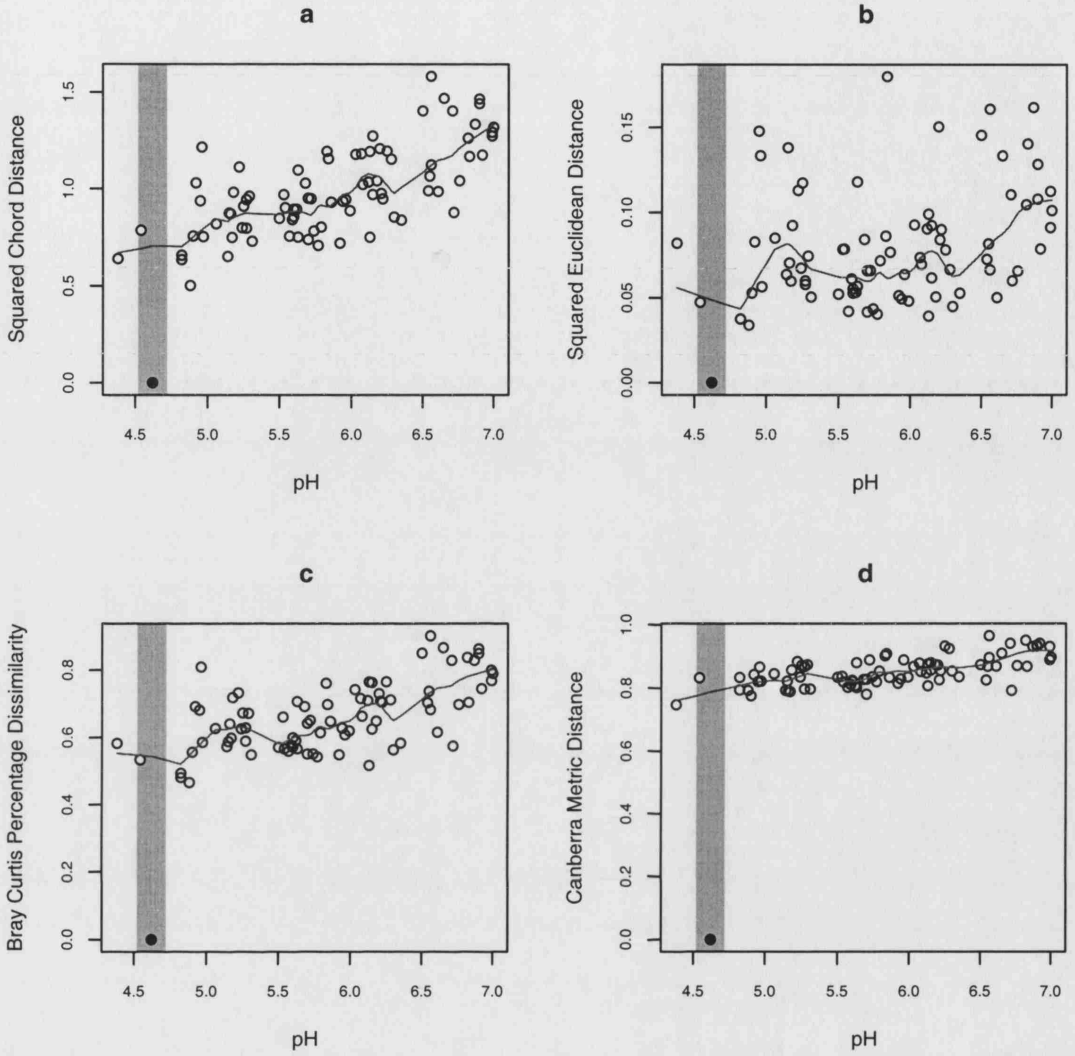


Figure 74: Scatterplot showing the dissimilarity between Loch Dallas and every other sample in the modern training set against pH. The trend line is a LOWESS smoother (span=0.2), and the shaded rectangle indicates the area ±0.1 pH about the pH of Loch Dallas. The 4 dissimilarity coefficients are presented; a) squared chord distance, b) squared Euclidean distance, c) Bray Curtis percentage dissimilarity and d) Canberra metric.

5.3.7.1 Performance of resemblance measures along an acidity gradient

To investigate how well various resemblance measures can discriminate amongst dissimilar and similar samples from the modern training set, five samples (DALL, FHIO, ACH, LOCH, BRAC) were selected that lie along the pH gradient in the 83-lake diatom and cladoceran training set. The resemblance between each of these five samples and the other samples in the training set was calculated for four resemblance coefficients; squared chord distance, squared Euclidean distance, Bray Curtis percentage dissimilarity and the Canberra metric (corrected for double zeroes). Scatterplots of pH against resemblance illustrate where along the gradient the most similar samples occur. These are shown in Figure 74-Figure 78.

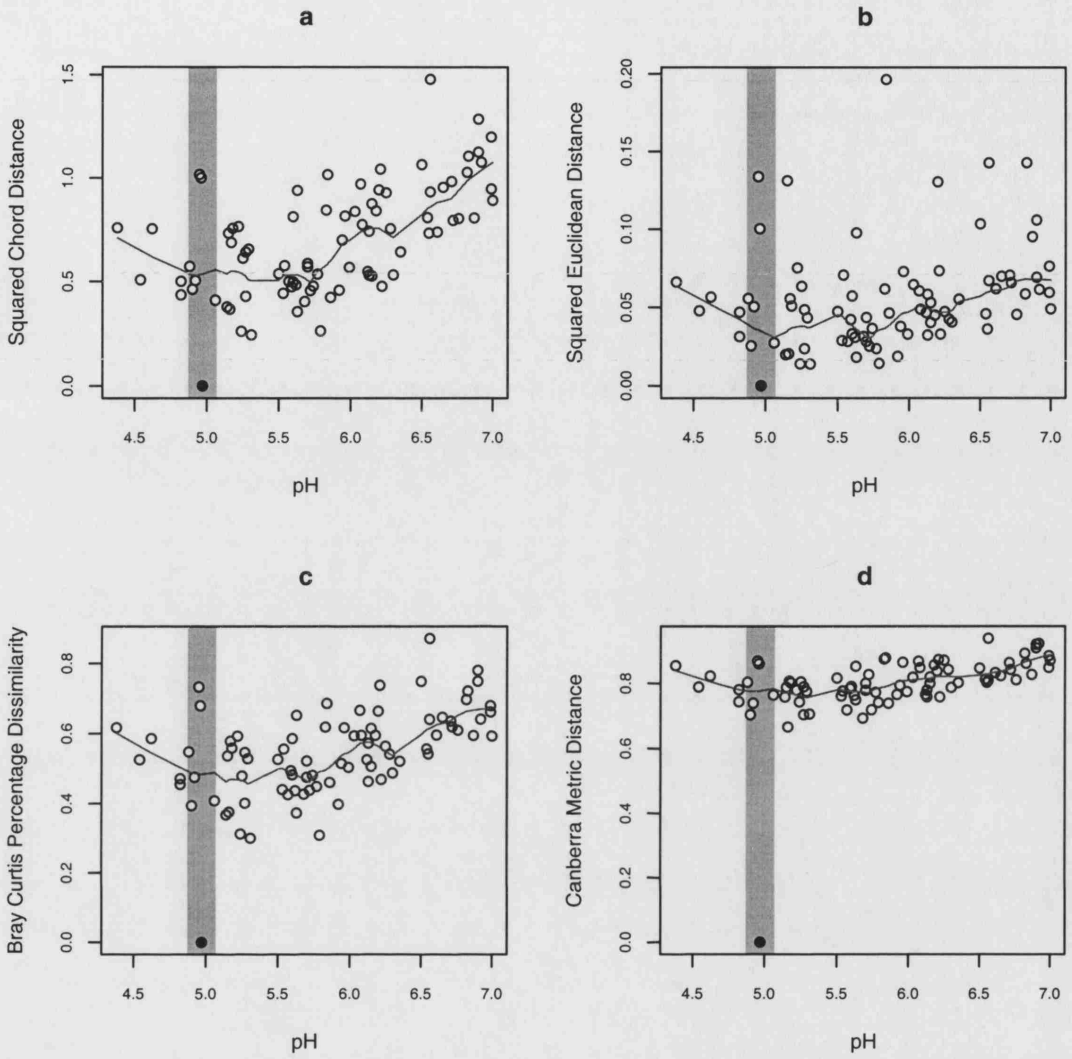


Figure 75: Scatterplot showing the dissimilarity between Lochan Fhionnlaidh and every other sample in the modern training set against pH. The trend line is a LOWESS smoother (span=0.2), and the shaded rectangle indicates the area ±0.1 pH about the pH of Loch Fhionnlaidh. The 4 dissimilarity coefficients are presented; a) squared chord distance, b) squared Euclidean distance, c) Bray Curtis percentage dissimilarity and d) Canberra metric.

It is clear from the diagrams in Figure 74-Figure 78 that both the Canberra metric and the squared Euclidean distance have difficulty is discriminating between similar and dissimilar samples along the pH gradient. The resemblance values indicated by the Canberra are consistently high for the five samples along the gradient indicating a high level of dissimilarity. The short range in minimum and maximum values for the Canberra metric also indicates that this measure is not a very sensitive measure of resemblance along the gradient.

The Bray Curtis percentage dissimilarity and the squared chord distance measures both show a strong trend in the plots along the gradient. The plotted trend lines for both these

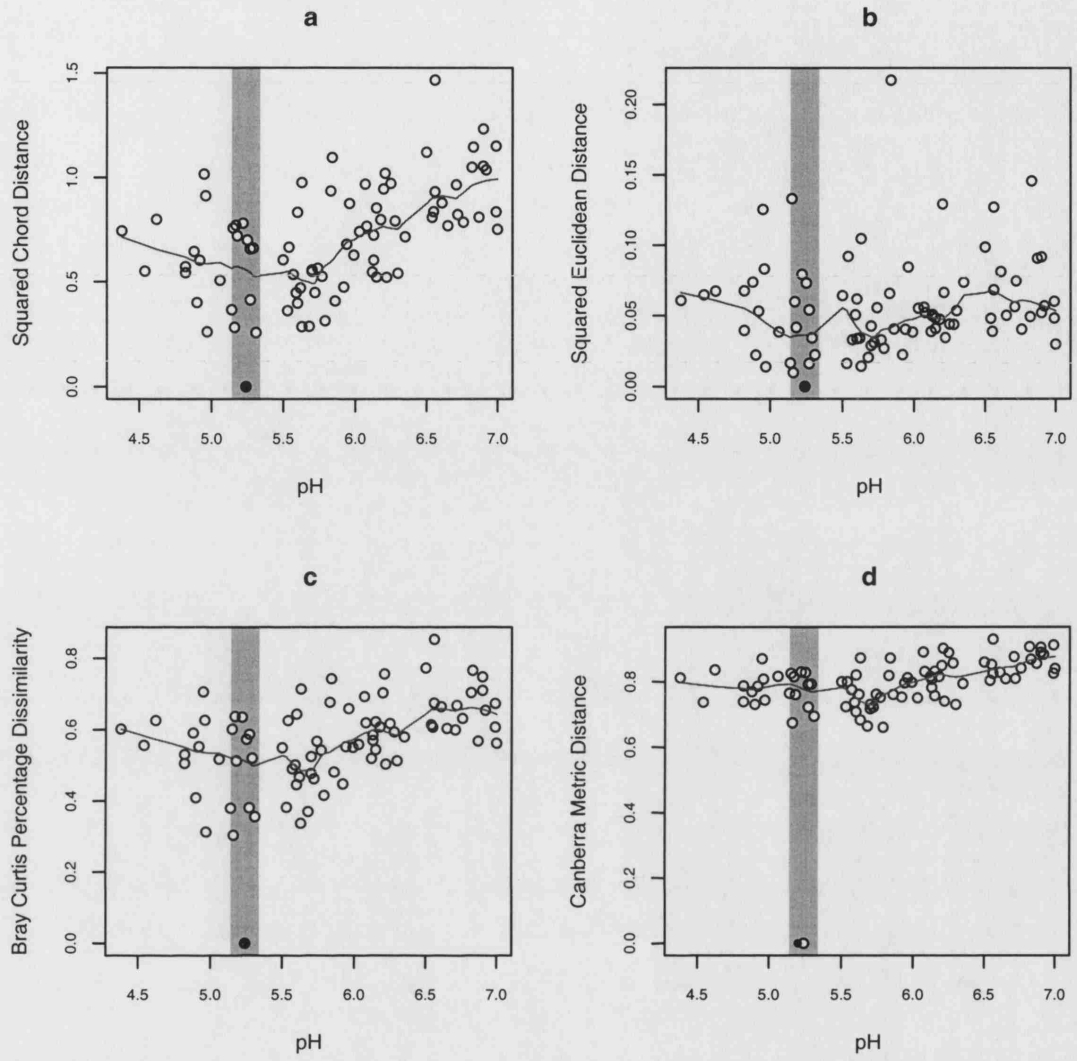


Figure 76: Scatterplot showing the dissimilarity between Loch na h'Achlaise and every other sample in the modern training set against pH. The trend line is a LOWESS smoother (span=0.2), and the shaded rectangle indicates the area ±0.1 pH about the pH of Loch na h'Achlaise. The 4 dissimilarity coefficients are presented; a) squared chord distance, b) squared Euclidean distance, c) Bray Curtis percentage dissimilarity and d) Canberra metric.

measures show the lowest dissimilarity values are seen between samples occurring at similar positions along the gradient. Furthermore, the trend line indicates that dissimilarity increases quickly as one moves along the gradient, away from the observed data point.

The squared chord distance results suggest that this method is the more sensitive measure of resemblance along the acidity gradient because there is a greater range between the minimum and the maximum resemblance values and the plotted trend line through the data shows a higher magnitude of change. About the centre of the gradient in particular, the trend lines are significantly curved, more so than any other method.

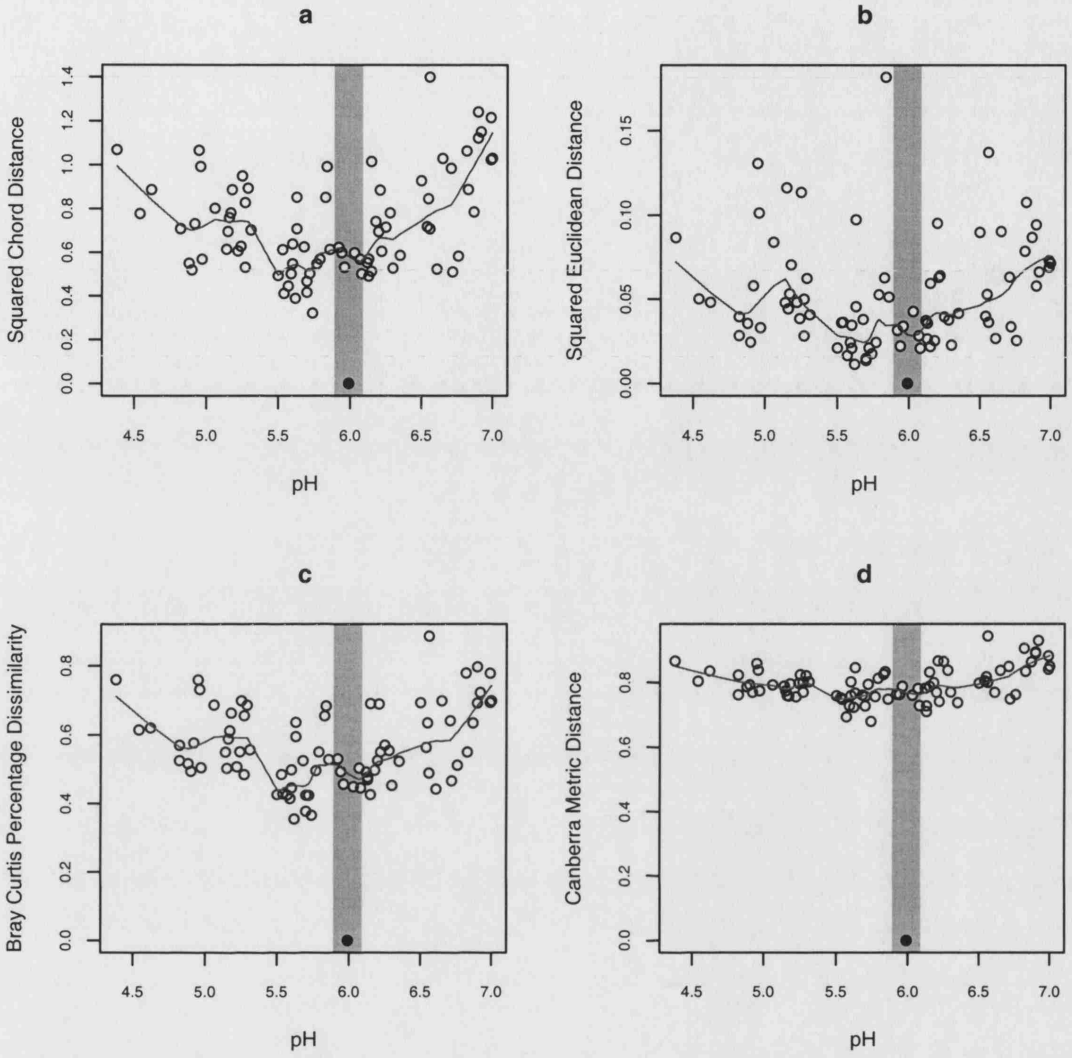


Figure 77: Scatterplot showing the dissimilarity between Loch Toll an Lochain and every other sample in the modern training set against pH. The trend line is a LOWESS smoother (span=0.2), and the shaded rectangle indicates the area ±0.1 pH about the pH of Loch Toll an Lochain. The 4 dissimilarity coefficients are presented; a) squared chord distance, b) squared Euclidean distance, c) Bray Curtis percentage dissimilarity and d) Canberra metric.

From these plots it would appear that the squared chord distance was best able to represent the known differences between the acidities of the lakes in the modern training set than the other measures.

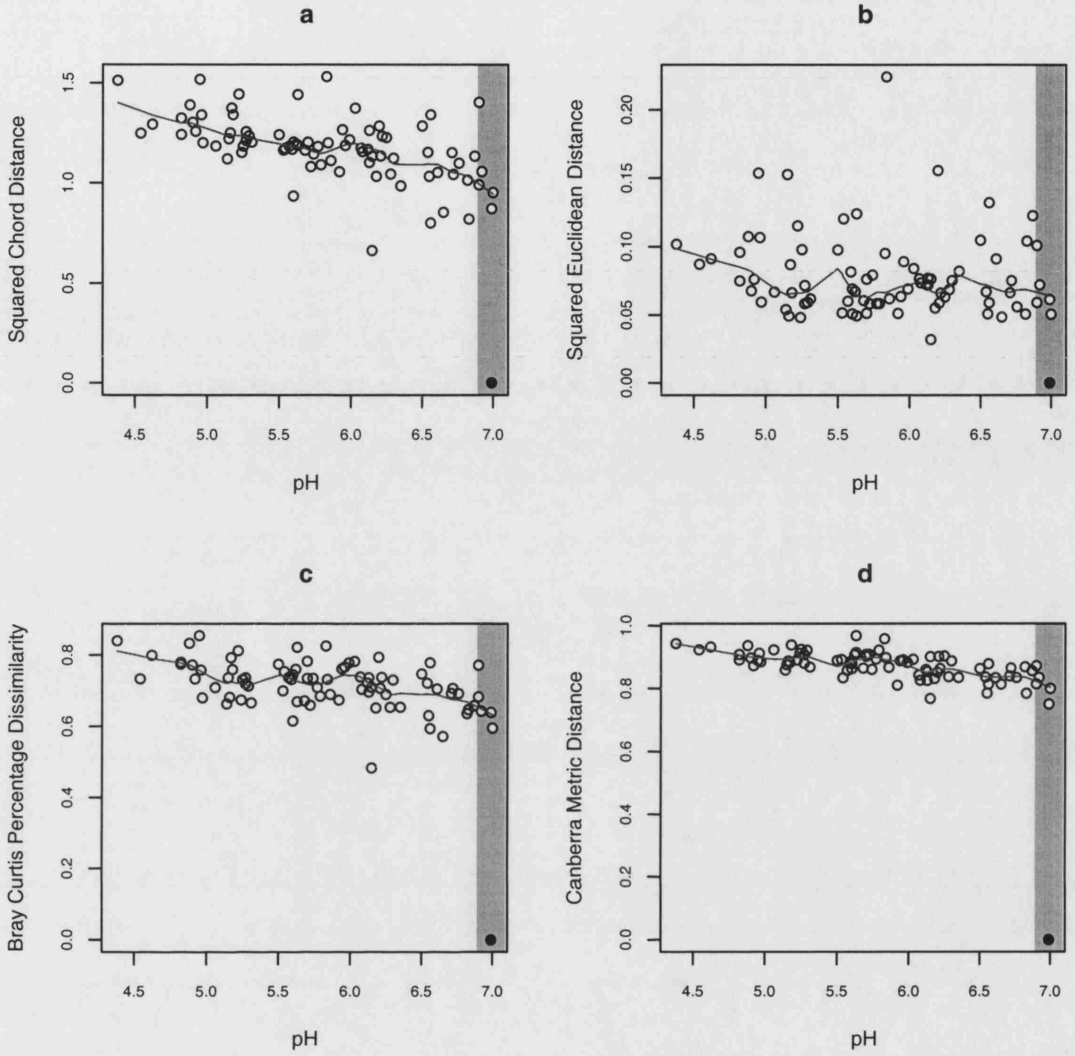


Figure 78: Scatterplot showing the dissimilarity between Loch nam Brac and every other sample in the modern training set against pH. The trend line is a LOWESS smoother (span=0.2), and the shaded rectangle indicates the area ±0.1 pH about the pH of Loch nam Brac. The 4 dissimilarity coefficients are presented; a) squared chord distance, b) squared Euclidean distance, c) Bray Curtis percentage dissimilarity and d) Canberra metric.

5.3.7.2 Determining critical values for the analogue matching data sets

In Section 5.3.7, a number of ways of generating critical values for resemblance measures were discussed. In this section I will use some of these methods to determine critical values for the 4 analogue matching data sets before going on to apply these data sets in an illustration of analogue matching for the lakes in the UKAWMN.

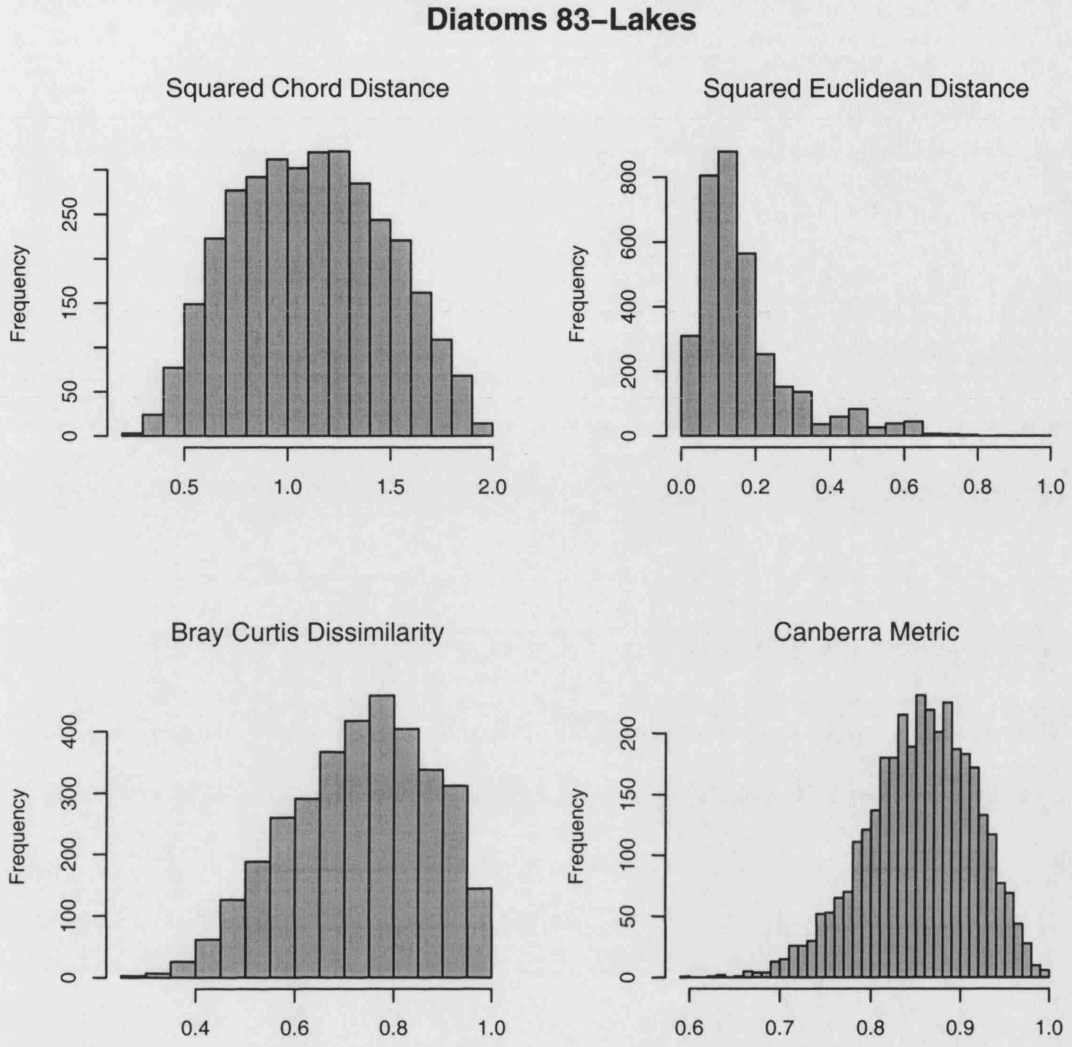


Figure 79: Histograms showing the distributions of resemblance values for the 83-lake diatom training set calculated for four dissimilarity coefficients; squared chord distance, squared Euclidean distance, Bray Curtis percentage similarity and Canberra metric. Bin width determined using Scott's rule (1979).

Of the four methods discussed above, I will restrict the majority of this section to just two of the methods described, but I will discuss the applicability of all four methods.

It would be difficult to establish arbitrary critical values for the four data sets using expert knowledge alone. The diatom data contained in the UKAWDDS have not been collated or analysed before now, so the properties of the data are not known. There has also been little work done on cladoceran community composition in acid sensitive upland lakes in the UK. So the limited amount of *a priori* knowledge would hamper any attempt to set critical values from these sources alone.

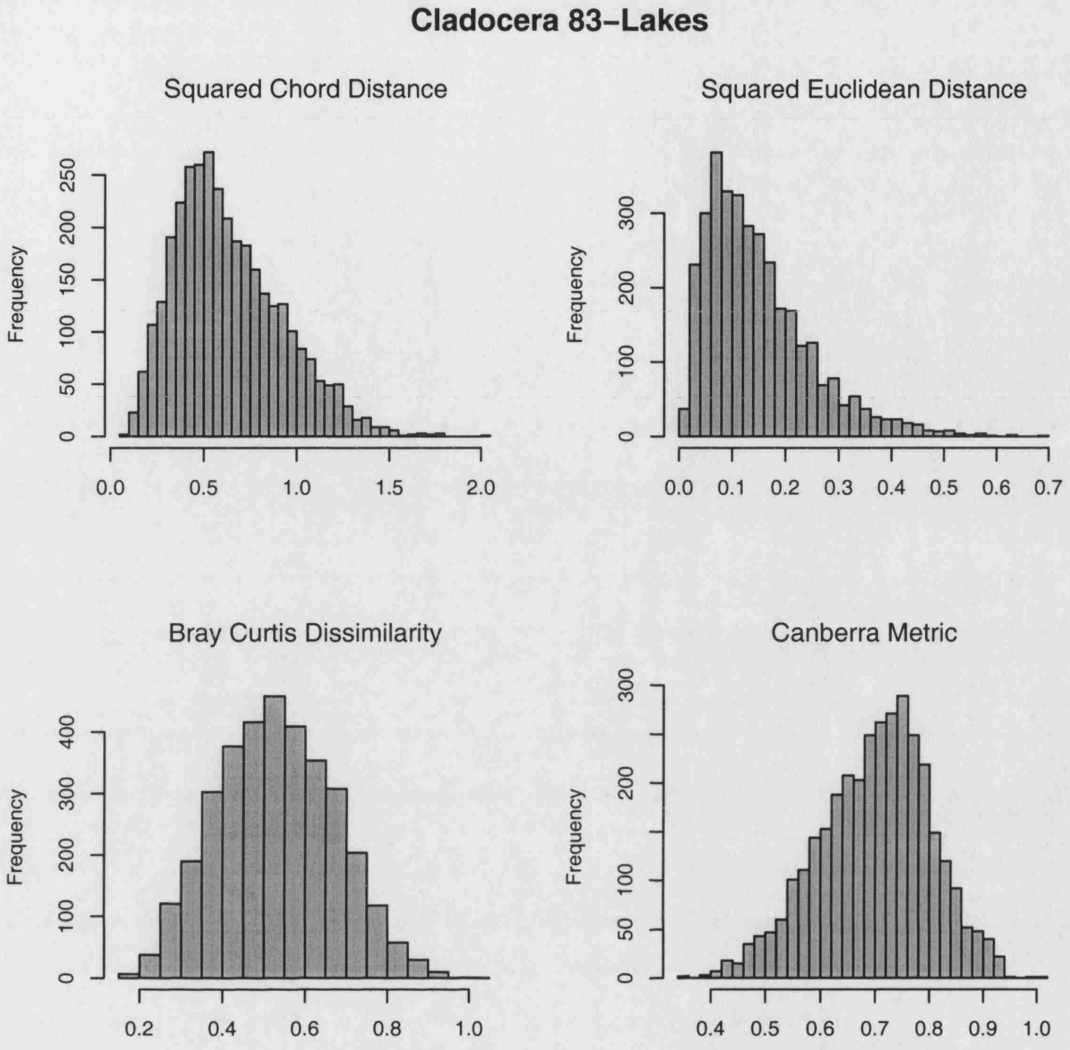


Figure 80: Histograms showing the distributions of resemblance values for the 83-lake Cladocera training set calculated for four dissimilarity coefficients; squared chord distance, squared Euclidean distance, Bray Curtis percentage similarity and Canberra metric. Bin width determined using Scott's rule (1979).

The method of determining the critical values empirically from the observed distribution of resemblance values looks, on the face of it, very appealing. It has the advantage over the *a priori* knowledge method in that it allows the data themselves to determine the critical values.

Figure 79-Figure 82 show the observed distributions of resemblance values for the four analogue matching data sets. Both the squared Euclidean distance and the Canberra metric show strongly skewed distributions of the dissimilarity values. The values are right skewed for the squared Euclidean distance indicating that the majority of the samples in the training sets are quite similar to each other, with only a few of the samples differing

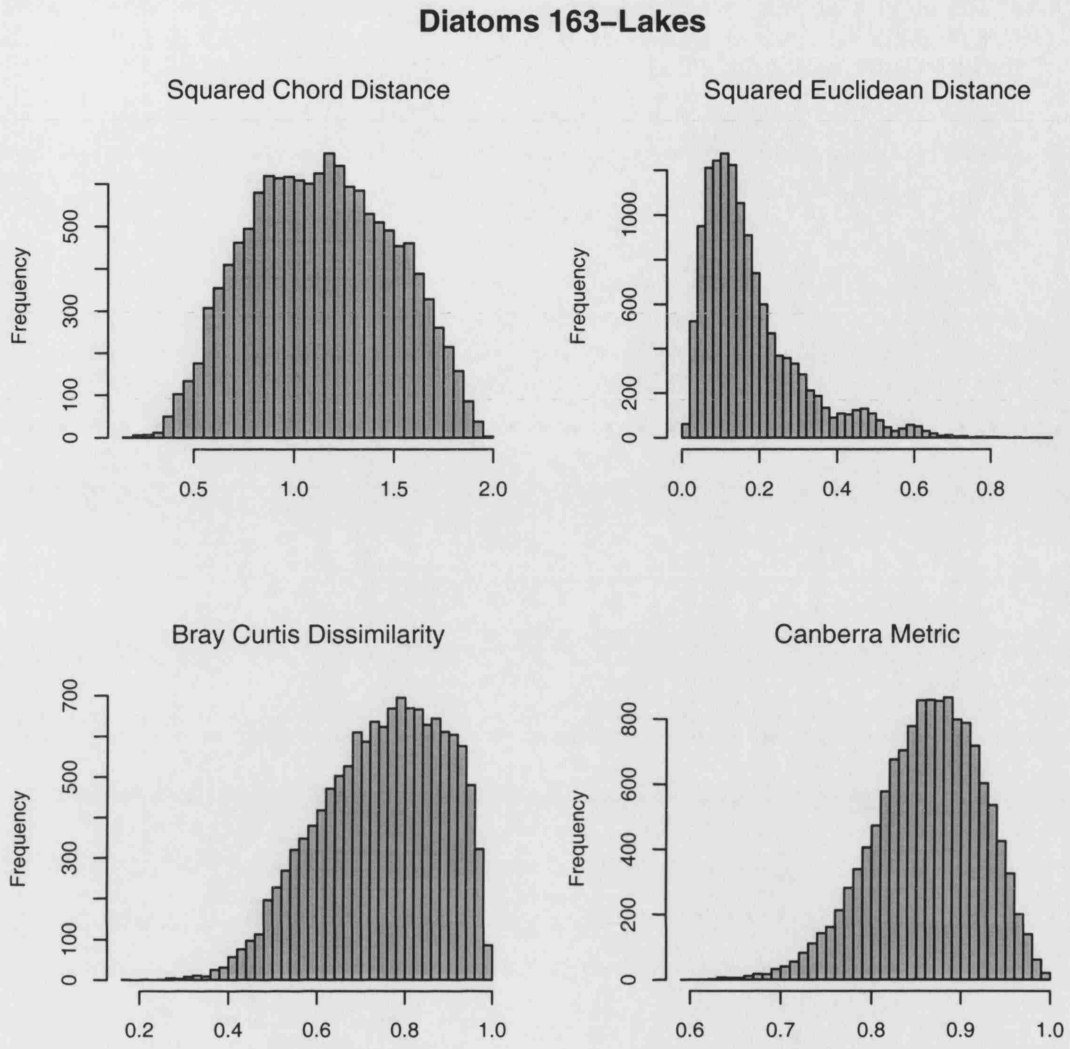


Figure 81: Histograms showing the distributions of resemblance values for the 163-lake diatom training set calculated for four dissimilarity coefficients; squared chord distance, squared Euclidean distance, Bray Curtis percentage similarity and Canberra metric. Bin width determined using Scott's rule (1979).

significantly from each other. Critical values for this distribution would be extremely low, and for samples to be classed as close modern analogues they would have to be extremely similar in their species composition. The reverse is true of the Canberra metric. The distribution of the values for this resemblance measure is left skewed with relatively few low values. Critical values for this coefficient would be correspondingly high with the result that many of the analogues selected for a sample would be quite different to the modern sample in terms of their species composition. Table 41 shows the critical values for each of the four resemblance measures plotted in Figure 79-Figure 84.

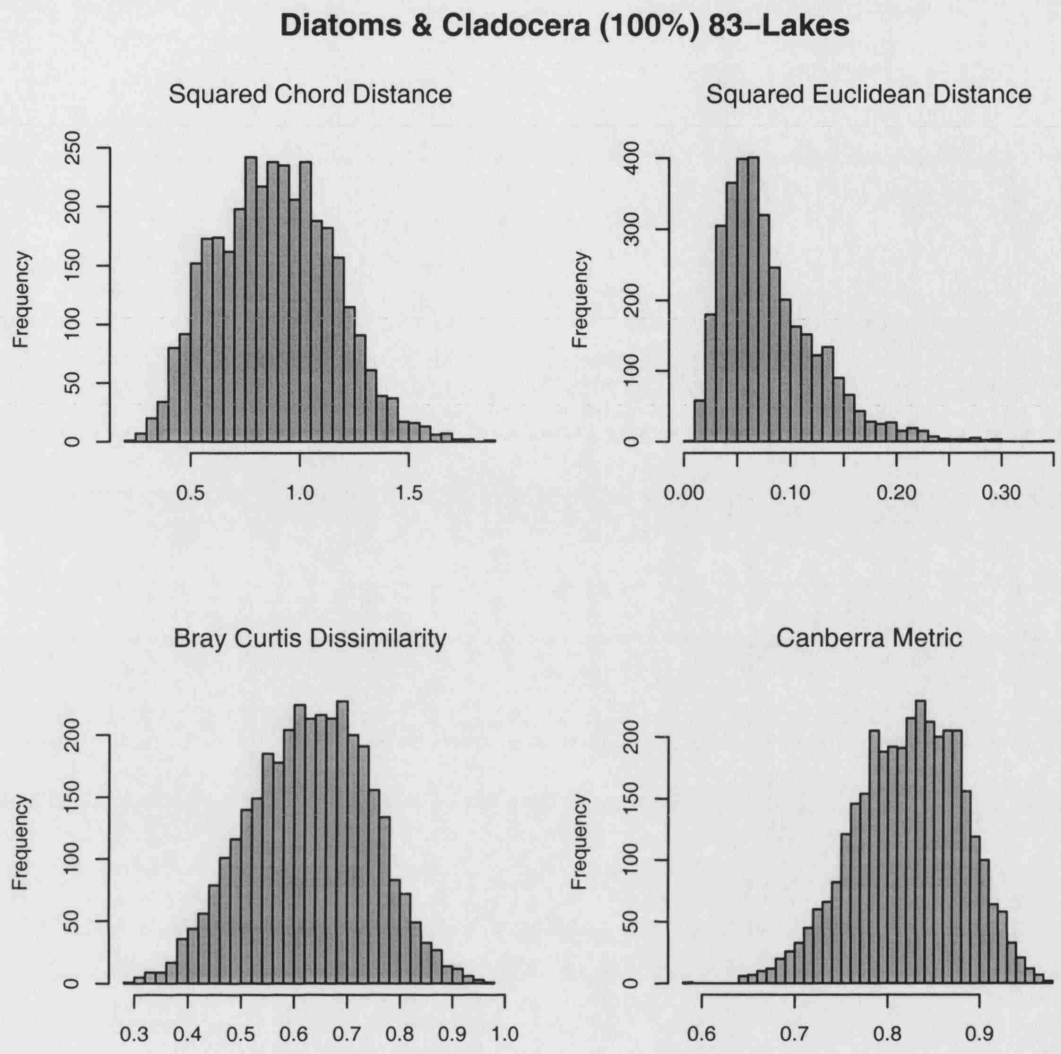


Figure 82: Histograms showing the distributions of resemblance values for the 83-lake diatom and Cladocera training set calculated for four dissimilarity coefficients; squared chord distance, squared Euclidean distance, Bray Curtis percentage similarity and Canberra metric. Bin width determined using Scott's rule (1979).

Across the four datasets, the distributions of the Bray Curtis percentage dissimilarity values range from being slightly left skewed to approximating a normal distribution. Though the skewness is much less extreme than in either the Canberra metric or the squared Euclidean distance, being left skewed again increases the level of dissimilarity allowed before the critical value of the coefficient is reached.

The distributions of the observers values of the squared chord distance coefficient across the four data sets is, in general, less skewed than either of the other three resemblance measures, Only for the Cladocera data set does the distribution show a significant deviation from the normal, resulting in a right skewed distribution. Section 4.2 above

showed that there is only a small degree of difference in species composition across the range of samples from upland, acid sensitive lakes. This is reflected in the distribution of the resemblance values for the squared chord distance coefficient for the dataset. In this case, it is likely that the deviation from the normal curve reflects the close degree of similarity of the cladoceran fauna between many of the samples in the training set.

The distributions of the resemblance coefficients for the four training sets illustrate the major problem with establishing critical values empirically from the observed dissimilarities. As the critical values are determined by the percentiles of the distribution of the dissimilarity values across the modern training set, they are greatly influenced by deviations from a normal distribution.

This is particularly important in cases where the distribution of dissimilarity vales is strongly left skewed. In such situations many more analogues would be identified than actually occur. These analogues would be quite dissimilar in their species composition because higher dissimilarities were chosen for the critical values. This is less important where the distribution is right skewed because even though fewer analogues might be identified than actually exist, these analogues would at least have very similar species compositions to the fossil sample.

So, whilst the idea of deriving the critical values from the observed distribution of resemblance values is very appealing, one must carefully check for deviations from the normal distribution in the observed dissimilarities.

It is clear from the discussion above, that more-robust methods of establishing the critical values used to select modern analogues for fossil assemblages are required. A simple alternative would be to generate pseudo-random numbers that follow a normal distribution. This has been performed for each of the four training sets using the R statistical software for the squared chord distance coefficient.

First, a training set of $\frac{1}{2}N(N-1)$ randomly generated normal deviates is produced. Values for this training set range from -1 to 1. The values are then rescaled such that they range

from 0 to 2, the minimum and maximum values taken by the squared chord coefficient for perfectly similar and perfectly dissimilar samples respectively.

The result is a training set of size $\frac{1}{2}N(N-1)$ random resemblance measures in the range 0...2. One can then take the percentiles of this normally distributed dataset and use the values of the extreme 1st, 2.5th, 5th or 10th percentiles as the critical values for comparing modern and fossil samples.

To illustrate the method, a pseudo-random set of resemblance data ranging from 0...2 of size $\frac{1}{2}N(N-1)$, where N=83, was generated using the random normal deviate generator in R 1.5.1. (Ihaka and Gentleman 1996). The extreme 1st, 2.5th, 5th and 10th percentiles of this distribution were calculated using the quantile() function in R. Table 45 shows the critical values as derived from the pseudo-random data. Figure 83 shows the empirically derived critical values for the four analogue matching training sets as bars, with the pseudo-randomly generated critical values indicated by dashed lines intersecting the relevant group of bars.

The pseudo-random critical values are assumed to be those critical values that one would expect to find if the resemblance values followed a normal distribution. If one is going to use the extreme percentiles of the distribution of these resemblance values then the observed resemblance values must follow a normal distribution. This is because we are hypothesising that the population sample contained in our modern training set includes many samples that are somewhat dissimilar to each other but that only a few samples are very similar or very dissimilar. At the upper extreme we expect to find those samples that are very dissimilar from the other modern samples, these are expected to be relatively rare, especially if the modern training set contains samples from a similar environment. At the lower extreme we expect to find few samples that are very similar to each other. If these assumptions are met, then we can infer that the similarity we would expect to occur between two similar samples is that level of similarity observed at the extreme lower end of the distribution of resemblance values. Deviations from the normal distribution lead to overly large or small critical values that would indicate that greater or fewer numbers of samples are similar to one another than one would expect to occur.

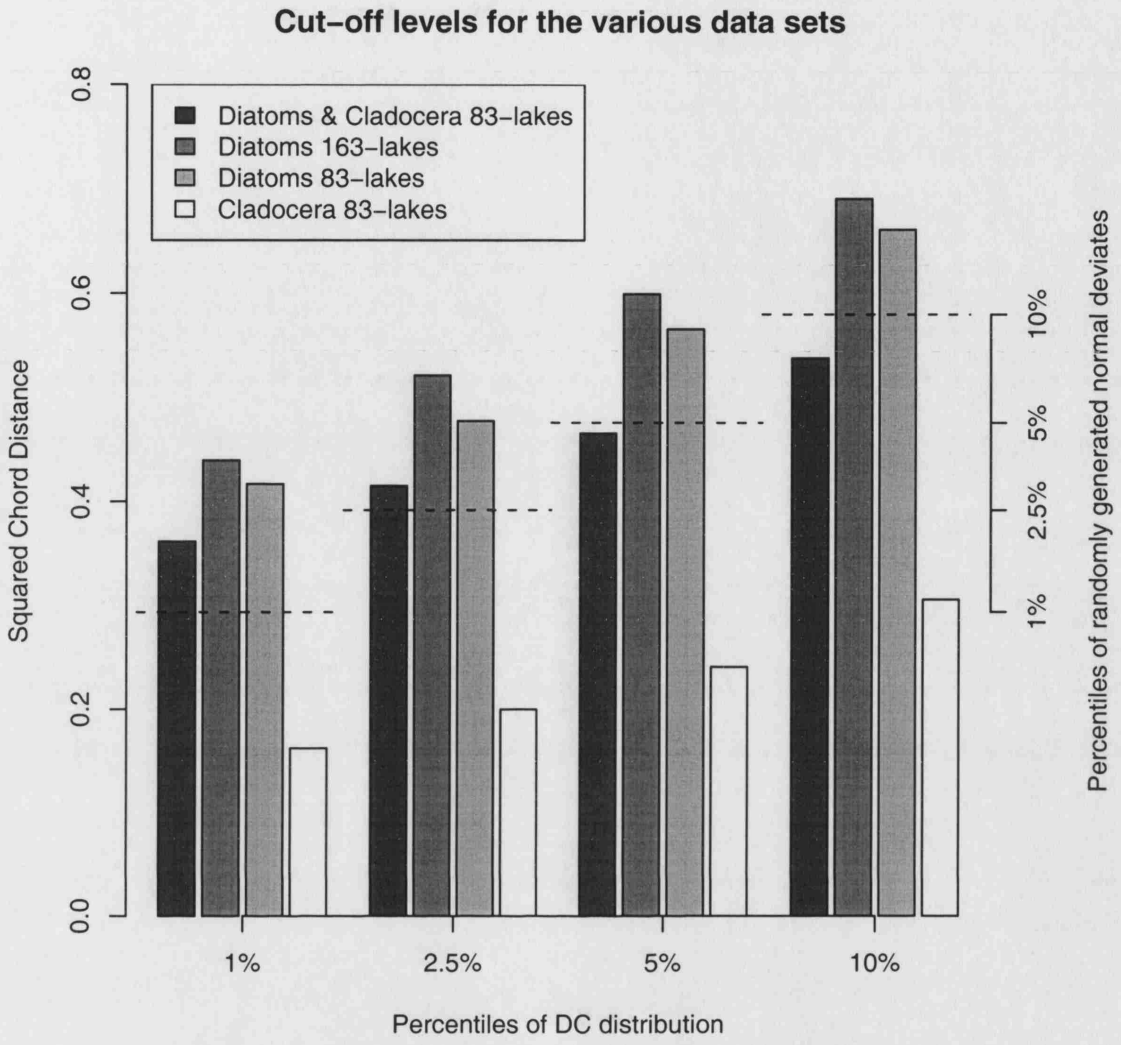


Figure 83: Barplot comparing the empirically-derived critical values of the squared chord distance coefficient with the critical values of randomly derived data. The bars show the critical values for the 1st, 2.5th, 5th and 10th percentiles of the observed distribution of resemblance measures in the training sets discussed. The dotted lines indicate the critical values for the same percentiles of random data.

The empirically derived critical values for the cladoceran training set are all significantly lower than the critical values implied by the pseudo-random data set. This reflects the skewed distribution of the squared chord distance values as shown in Figure 80.

The empirically derived critical values for the two diatom training sets are consistently higher than those derived from the pseudo-random data. The critical values for the 163-lake training set are also consistently greater than those for the 83-lake training set. The discrepancy between the empirically-derived critical values for the diatom only training sets is generally of the order of 0.1, approximately allowing 5% more dissimilarity for samples to be classed as analogues. The difference between the two sets of critical values is

significantly less than the difference between the two sets of critical values for the cladoceran training set. This pattern is consistent with the left skew pattern shown in the distribution of the squared chord distance data for both the 83-lake and the 163-lake training sets, as seen in Figure 79-Figure 82.

The empirically-derived critical values for the diatom and cladoceran training set follow much more closely the values suggested by the pseudo-random data. The 1st percentile critical value is *ca.* 0.075 greater than that suggested by the pseudo random data, whilst the 2.5th and 5th percentile values are very similar indeed. The values for the 10th percentile also differ slightly, with the observed critical value being slightly lower than that established using the pseudo-random data.

The main advantage over the empirically-derived critical values is that this method is independent of the observed data, and as such is not limited by deviation from the normal distribution that can affect considerably the empirically-derived critical values.

5.3.7.3 Performance of the squared chord distance coefficient along the main gradients in the diatom and cladoceran data.

In the previous section the squared chord distance coefficient was shown to perform consistently well at discriminating samples known to be similar from those known to be dissimilar along the pH gradient. Critical values for the squared chord distributions have been established using two methods; one based on the observed distribution of resemblance values between each sample and every other sample in the modern training sets, the other based on pseudo-random normal deviates generated on the scale of the squared chord distance.

A redundancy analysis (RDA) of the 83-lake diatom and cladoceran training set was performed. The redundancy analysis extracts the main patterns of variation in the species data with the constraint that these patterns are linear combinations of the observed environmental data. In the simple sense, RDA extracts those patterns that can be explained by the measured data. The RDA itself is discussed more fully in the following section (Section 5.2 above). The site scores along the first two of axes of the RDA represent a theoretical response gradient that is some combination of our measured environmental

data. So now, instead of the gradient being simply that of a single variable in the data set, it is a combination of variables that best explain the distributions of the species found in the samples along the gradient.

We can take these site scores along the first or second RDA axis and plot the dissimilarity between a given sample and every other sample in the training set. By repeating this procedure for samples that fall at various distances along this theoretical gradient we can assess where along the gradient suitable analogues occur, in much the same way as was performed in Section 5.3.7.1 above. The difference is that we are taking into account floristic and faunistic differences between the samples in the training set when attempting to determine where along the gradient suitable analogues occur. It should be noted that this is restricted to the modern training set alone and does not indicate where along a gradient we might have good modern analogues for *fossil samples*; however, it does indicate where along the gradient we might have suitable coverage of the environmental space so that modern analogues might be identified for fossil samples.

The first axis of the RDA of the diatom and the cladoceran data is mainly an acidity gradient being closely correlated with pH, calcium, alkalinity etc. However, other important variables in the training set are accounted for by this first axis, namely conductivity (including magnesium, potassium, sodium and chloride), lake depth (MLDepth), altitude (LAlt and MaxAlt), TOC and aluminium.

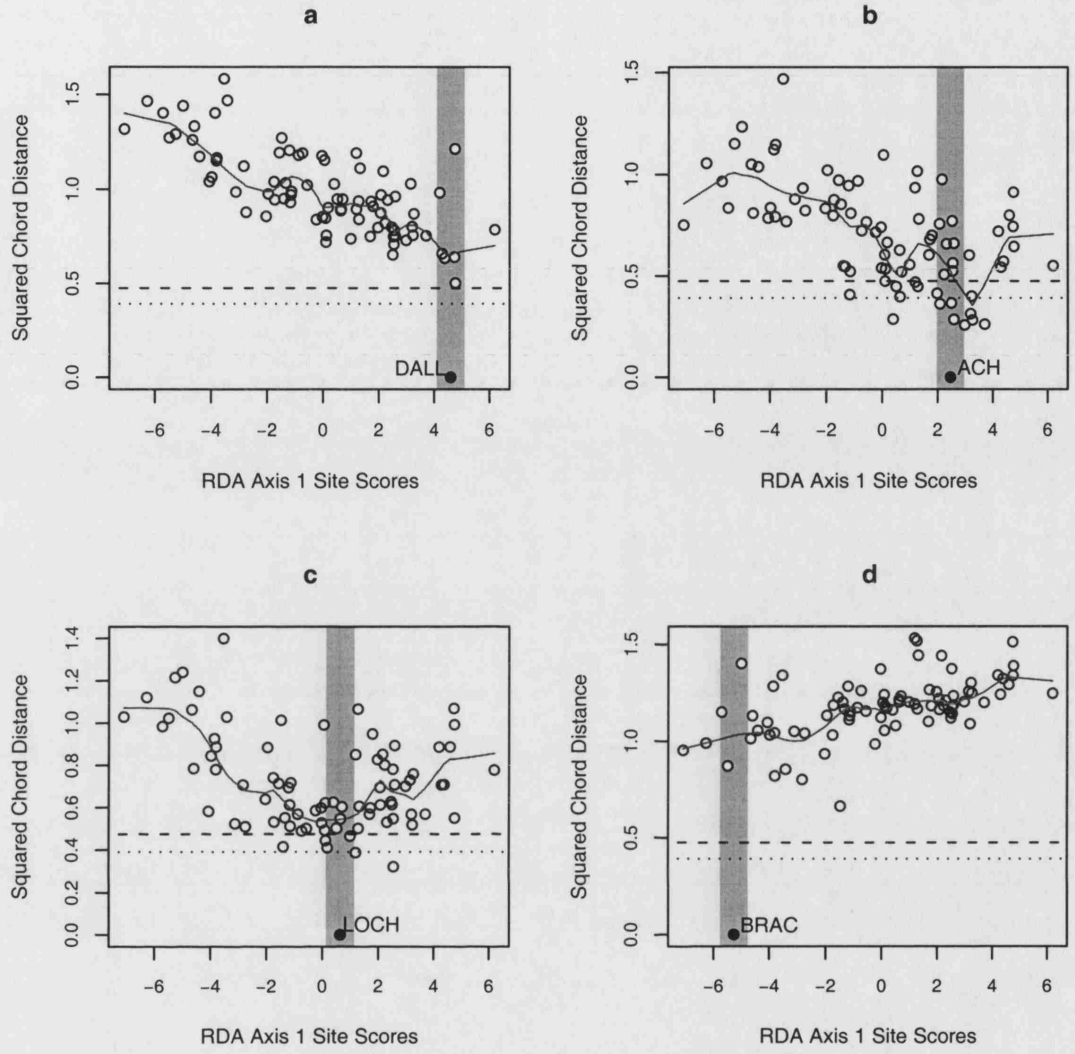


Figure 84: Scatterplot showing the squared chord dissimilarity between each named lake and every other sample in the modern training set against the RDA axis 1 site scores for each site in the training set. The trend line is a LOWESS smoother (span=0.2). 4 lochs are presented; a) Loch Dallas, b) Loch na h'Achlaise, c) Loch Toll an Lochain and d) Loch nam Brac. The grey bar is presented as an aid for interpretation and simply illustrates the point on the gradient of the named lake.

Figure 84 shows gradient dissimilarity plots for four samples (DALL, ACH, LOCH and BRAC) along the first RDA axis. The data points represent the dissimilarity between the given site and every other site in the modern training set, ordered by their position on the theoretical gradient. The main pattern in the data points is shown by a LOWESS smoother (span=0.25), and the critical values of the 2.5th and 5th percentiles of the pseudo-random data are indicated by the dotted and dashed horizontal lines respectively. The grey, vertical bar indicates the region of the gradient where analogues would be expected.

A similar pattern to that shown for the pH gradient in the training set is apparent in Figure 84; namely that samples located towards the centre of the gradient (ACH and LOCH) have a

number of close analogues in the modern training set. At the more extreme ends of the gradient, however, no analogues that meet the critical values for the 2.5th and 5th percentiles are found. The LOWESS smoothers again illustrate that the squared chord distance coefficient takes lower values for those samples known to be similar to each other in terms of the floristic and faunistic relationships with the environment. It is further apparent that towards the extremes of the gradient fewer samples exist in the training set and it is intuitive that fewer (if any) analogues are likely to be found from such a small sample of the environmental space.

The results presented in Section 5.3.7 provide the basis for applying the analogue matching training sets to the task of identifying modern analogues for the lakes of the UKAWMN. Critical values have been established that allow one to determine whether a particular sample is sufficiently similar to the fossil sample to be considered as an analogue. The squared chord distance coefficient has also been shown to have both desirable theoretical properties for the types of data being used, but that the coefficient also performs well in practical situations when compared to examples from the other major groups of resemblance coefficients (simple and equal weight coefficients), or those that are popular in numerical ecology (Bray Curtis percentage dissimilarity).

5.4 Applied analogue matching

The previous sections of this chapter have established a number of factors required for the analogue matching approach to be applied to lake sediment cores. Critical values for determining the degree of resemblance required for samples to be identified as close analogues for another sample have been established using two methods. The squared chord distance coefficient has been identified as having the greatest ability to reflect the known similarity between samples in the modern training sets in the resemblance of the sub fossil diatom and cladoceran assemblages.

The previous section (see Section 5.2 above) has shown that the combined diatom and cladoceran data set contains two main patterns in the data that closely reflect an acidity and acid-sensitivity gradient and a complex gradient of physico-chemical properties, such as lake depth and altitude, and conductivity and calcium. These patterns are a combination of

the main patterns found in the diatom and cladoceran training sets respectively, as previously discussed (see Chapter 3 and Chapter 4).

Government policy on emissions reductions is in part driven by the benefit to ecosystem health that reductions in acid deposition would bring about. It is also essential that recovery in the light of these emission reductions is assessed with regard to progress made towards the pre-impact status of the disturbed surface waters. For these reasons there has been a growing need to establish reference conditions for acidified systems so that the amount of biological damage caused by acid deposition can be identified, and that the likely ecological benefits of reducing emissions can be weighed against the cost of implementing those same reductions.

Analogue matching is a method by which reference conditions can be established. Analogue matching differs from the traditional palaeolimnological approach where transfer functions have been the norm for reconstructing pH, for example, from the known response of the biology to the chemical parameter one is trying to reconstruct. The transfer function approach has two limitations, first it is generally limited to a single variable, and second, the end result is a reference condition based on a hydrochemical criterion. Analogue matching, however, is not constrained to a single variable and the reference state is biologically based; being established from the biological components measured at the modern analogue sites today.

To apply the analogue matching approach to core samples the lakes of the United Kingdom Acid Waters Monitoring Network (UKAWMN) were selected. These lakes are routinely monitored for signs of chemical and biological recovery. Establishing ecological reference conditions for these lakes is a significant and practical exercise that will inform governmental policy in ongoing discussions in Europe.

5.4.1 The United Kingdom Acid Waters Monitoring Network

The United Kingdom Acid Waters Monitoring Network (UKAWMN) was established in 1988 (Patrick, Battarbee, and Jenkins 1996) to assess the response of the aquatic environment to reductions in acid deposition at a range of sensitive sites across the UK. The network is funded by the Department of Environment, Food and Rural Affairs

(DEFRA) and the Department of Environment, Northern Ireland, and is maintained by ENSIS Ltd. The central aim of the UKAWMN is the detection of chemical *and* biological recovery from surface water acidification in the UK. 11 lake and 11 stream sites make up the network. Details of the sites are shown in Table 44 and a map showing their locations in Figure 86. Of these twenty two sites, only the eleven lake sites will be discussed further in this thesis.

The UKAWMN sites are typical examples of the more acid sensitive surface waters found in the UK, located on weathering-resistant bedrock that generate little acid neutralising capacity (ANC) to buffer the acid inputs. Sediment core analysis and the application of a diatom-pH transfer function has shown that most of the lake sites have acidified in the last 150 years or so (Battarbee *et al.* 1988a; Patrick *et al.* 1996). The geographical spread of the sites (Figure 86) encompasses a wide range of latitudes, altitudes and levels of incident acid deposition. Forested and non-forested catchments are also incorporated to provide a comparison between trends in land-use (e.g. Loch Chon and Loch Tinker).

Sampling at the UKAWMN includes quarterly water chemistry sampling for a range of determinands and a series of biological assays for epilithic diatoms, aquatic macrophytes and macroinvertebrates, and salmonid fish. Analysis of sediment trap samples for the remains of diatoms is also being undertaken to provide yearly monitoring of the total diatom population in the lakes.

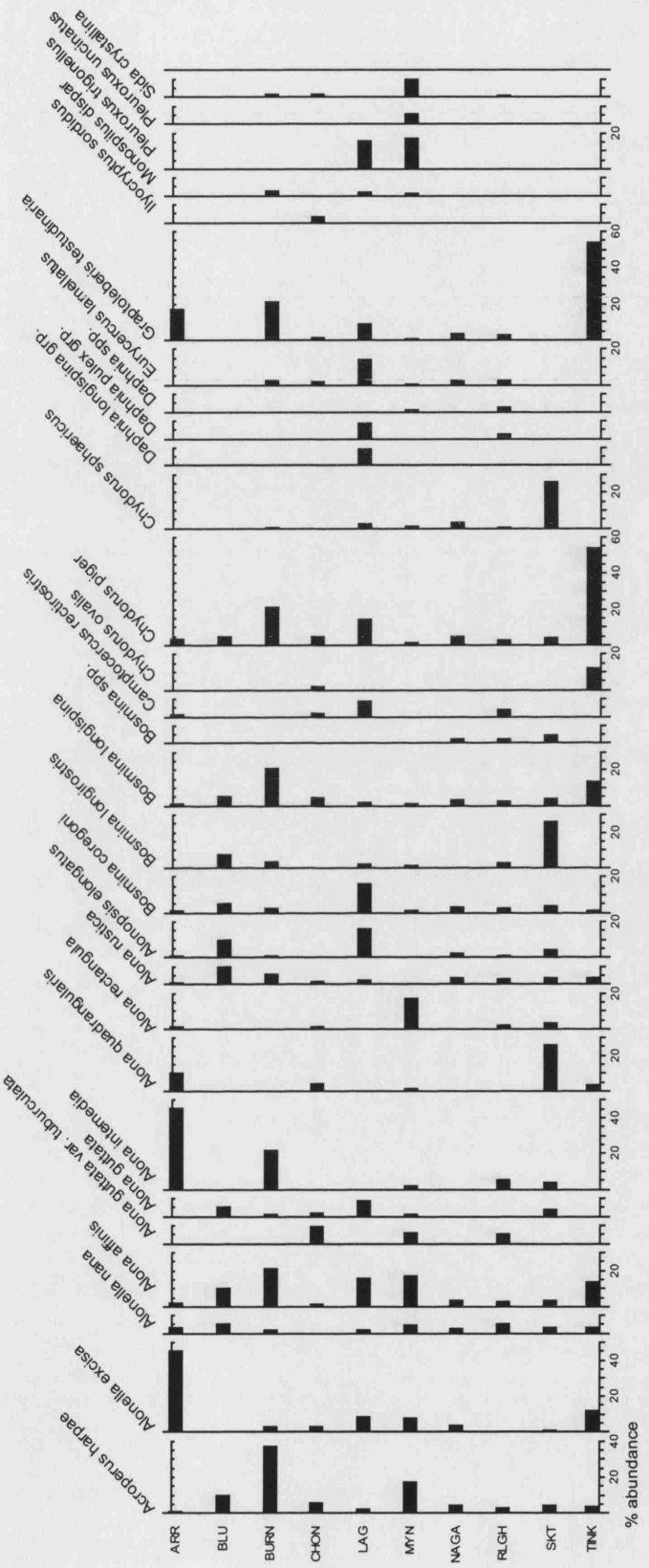


Figure 85: Cladoceran percentage abundance data for the pre-disturbance samples from the ten UKAWMN lakes for which analogue matching has been performed. See text for details of sediment sample depths and associated dates.

The lakes of the UKAWMN were selected for analogue matching for a variety of reasons. The lakes are well known palaeoecologically, having been sampled on a number of occasions for a range of proxies (diatoms, pollution indicators, heavy metals) and have good chronologies dated by the ^{210}Pb and SCP methods. Furthermore, sufficient suitable

Table 43: Sediment sample depth and ages for the ten UKAWMN pre-acidification samples. These data are taken from Patrick et al. (1995).

Site Code	Site Name	Sample Depth	Approx. Sample Age
ARR	Loch Coire nan Arr	15 – 16 cm	~ 1800 AD
NAGA	Lochnagar	17 – 18 cm	~ 1850 AD
CHON	Loch Chon	14 – 15 cm	~ 1850 AD
TINK	Loch Tinker	19 – 20 cm	~ 1820 AD
RLGH	Round Loch of Glenhead	21 – 22 cm	~ 1860 AD
SCOATT	Scoat Tarn	21 – 22 cm	~ 1800 AD
BURNMT	Burnmoor Tarn	23 – 24 cm	1906 AD
LAG	Llyn Llagi	22 – 23 cm	~ 1830 AD
MYN	Llyn cwm Mynach	22 – 23 cm	~ 1750 AD
BLU	Blue Lough	15 – 16 cm	~ 1750 AD

sediment was available from the pre-industrial period of the sedimentary record to allow for sub-fossil cladoceran analysis to be carried out.

The lakes in the UKAWMN are important sites for monitoring acidified systems and quantifying the impact of emission reduction policies throughout Europe. As such, it is important to establish baseline or reference conditions for these lakes so that monitored recovery can be placed in the proper context of the amount of damage caused by acid deposition and therefore allow one to evaluate progress made towards recovery of these ecosystems. These reference conditions may also be used as an idealised target for recovery and the cost implications of implementing new emissions reduction policies can be judged against the expected gains in biological recovery.

Table 44: Sites in the United Kingdom Acid Waters Monitoring Network (from Patrick et al. (1995)

Site Number	Site Name	Grid Ref	Lake / Stream
1	Loch Coire nan Arr	NG 808422	Lake
2	Allt a' Mharcaidh	NH 881045	Stream
3	Allt na Coire nan Con	NM 793688	Stream
4	Lochnagar	NO 252859	Lake
5	Loch Chon	NN 421051	Lake
6	Loch Tinker	NN 445068	Lake
7	Round Loch of Glenhead	NX 450804	Lake
8	Loch Grannoch	NX 542700	Lake
9	Dargall Lane	NX 449786	Stream
10	Scoat Tarn	NY 159104	Lake
11	Burnmoor Tarn	NY 184043	Lake
12	River Etherow	SK 116996	Stream
13	Old Lodge	TQ 456294	Stream
14	Narrator Brook	SX 568692	Stream
15	Llyn Llagi	SH 649483	Lake
16	Llyn cwm Mynach	SH 678238	Lake
17	Afon Hafren	SN 844876	Stream
18a	Nant y Gronwen	SN771556	Stream
18b	Afon Gwy	SN 824854	Stream
19	Beagh's Burn	D 173297	Stream
20	Bencrom River	J 304245	Stream
21	Blue Lough	J 327252	Lake
22	Coneyglen Burn	H 640885	Stream

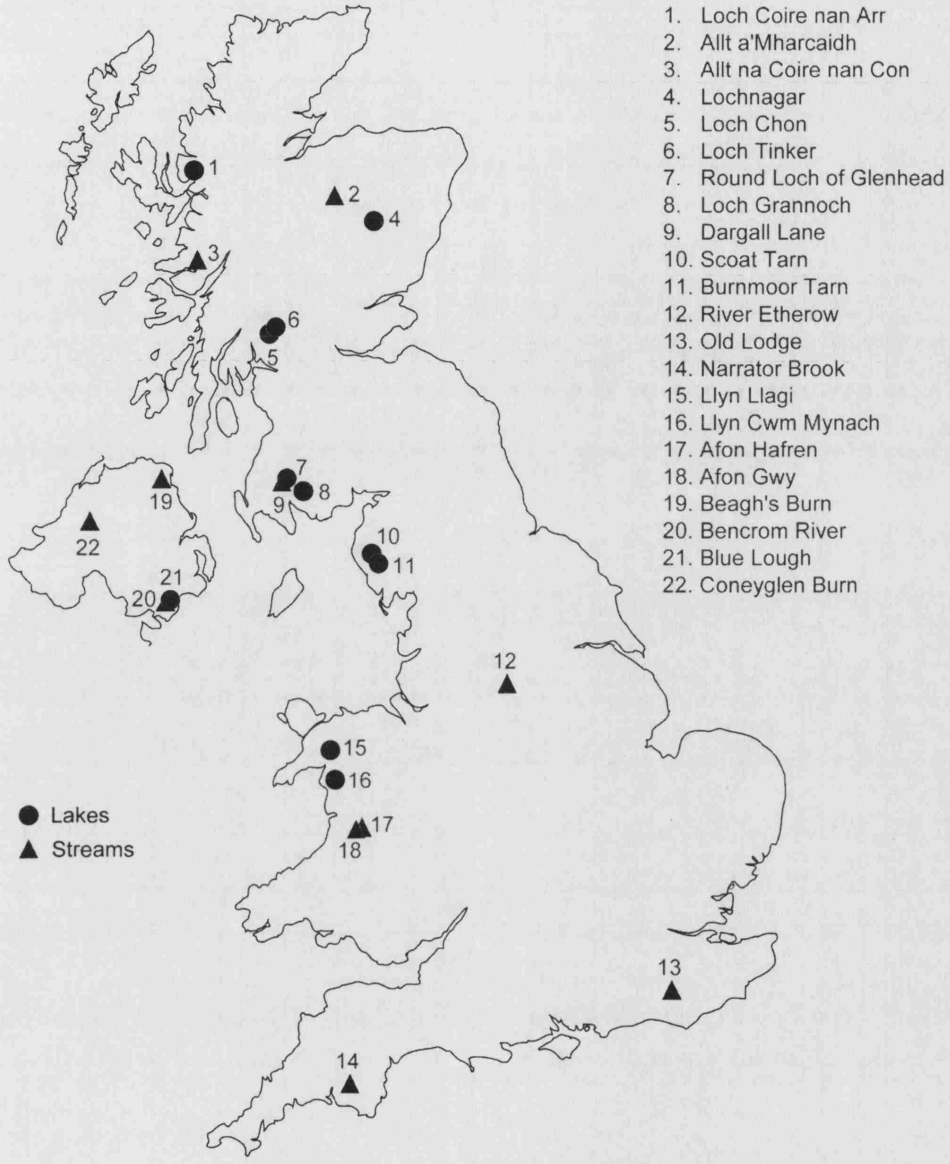


Figure 86: Map of the UK showing the location of the 22 sites in the United Kingdom Acid Waters Monitoring Network (source Patrick et al. *(1995)*

Loch Grannoch has not been included in this analysis because the core taken for the UKAWMN was not available for analysis at the time. Future work will include the expansion of the technique to the pre-acidification samples of this loch.

5.4.2 Cladocera and diatom-based analogue matching

Sediment cores for each of the UKAWMN lakes were located from the sediment archive of the Environmental Change Research Centre, from where the UKAWMN is administered by

ENSIS Ltd. The sediment cores are those taken in the early years of the network and are described in brief in Stevenson *et al.* (1991).

The diatom stratigraphy for these sediment cores has already established the pre-acidification period for the UKAWMN lakes in combination with pollution markers such as SCP's and heavy metals. Using the available data for each of the sediment records, the core bottom sample (See Table 43 for depths and ages of these sediment samples) from each sediment core was selected as reflecting pre-acidification conditions, and the diatom count data from this level were downloaded from AMPHORA, the ECRC's database. The corresponding sediment sample was then further prepared in the laboratory for cladoceran analysis using the standard method of deflocculation in warm 10% KOH (See section 2.2.2). The residue was plated and a sub-sample counted at 100 times magnification on a standard light microscope.

The diatom and cladoceran count data for each of the UKAWMN sediment samples were converted to percentage data based on the total sum of diatom and cladoceran individuals counted per sample. Expressing the data in such a way focuses the analysis on the species composition of the samples rather than the absolute abundance of species *per se.* Initially three analogue matching data sets were created using the full 163-lake diatom data set, the 83-lake cladoceran data set and the 83-lake Cladocera and diatom data set. In both the modern and the fossil data all taxa, irregardless of rarity, were included in the analysis.

The UKAWMN core bottom data were then appended to the diatom, cladoceran and diatom and cladoceran training set data files (as discussed in section 2.4.2) and the resulting files were processed using ANALOG version 1.6 (Line and Birks, unpublished computer program). ANALOG calculates the dissimilarity between each fossil sample and every sample in the modern training set using a given resemblance coefficient. For the analyses presented here the squared chord distance coefficient was used.

Following the procedure outlined in section 2.4.2, the output file from ANALOG was processed to identify those lakes from the modern training set that were more similar than the critical values derived from the 1st, 2.5th, 5th and 10th percentiles of the distribution of pseudo-random normal deviates.

The tables presented on the following pages show the results of a series of analogue matching procedures for each of the three analogue matching training sets. Unless otherwise stated, the tables distinguish between those lakes that have a similarity to a core bottom sample equal to or less than the critical values for the 5th and 10th percentiles respectively.

Subsequently, tables showing the range of chemistry values for the analogues selected are presented for the diatom and cladoceran training set only. In some instances the presented values are split into analogues that are at least as similar to the core bottom sample as the 5th percentile critical value and those that are at least as similar as the critical value for the 10th percentile would indicate.

The critical value for the 5th percentile used is 0.4753887 respectively. Table 45 below shows the complete list of critical values for the extreme percentiles of the distribution of the pseudo-random normal deviates.

Table 47-Table 54 also show the average values for each of the hydrochemical determinands across the selected modern analogues as well as the weighted average for each determinand based on the inverse of the dissimilarity between each modern sample and the fossil sample in question.

Table 45: Critical values for the squared chord distance coefficient based on the 1st, 2.5th, 5th and 10th percentiles of the distribution of randomly generated, normally distributed deviates of the range 0 to 2.

Percentile	Critical Value
1st	0.2936744
2.5th	0.3915955
5th	0.4753887
10th	0.5797419

Table 46: Number of "close" modern analogues identified for the pre-acidification sample from the 10-AWMN lakes for each of the training sets. The analogues are based on all taxa in the respective training sets that have maximum abundance greater than 2% in any one sample. The definition of "close" is based on the 5th percentile critical value as indicated in Table 45.

AWMN Lake	Diatoms 163	Diatoms 83	Cladocera	Diatoms & Cladocera
Loch Coire nan Arr	7	7	13	3
Blue Lough	0	0	0	1
Burnmoor Tarn	1	0	5	0
Loch Chon	6	4	0	0
Llyn Llagi	5	3	0	1
Llyn cwm Mynach	13	11	7	10
Lochnagar	0	0	33	2
Round Loch of Glenhead	13	9	2	9
Scoat Tarn	0	0	9	1
Loch Tinker	22	18	11	7

Table 47: Diatom and Cladocera-based modern analogues for Loch Coire nan Arr.

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	Toc	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
ARR	55.17	2.00	24.00	8.67	43.33	219.17	39.17	49.06	13.33	59.17	216.17	3.00	6.30	38.17	2.17	0.4297
CWEL	37.00	2.00	3.00	4.00	89.00	192.00	36.00	29.00	7.00	46.00	175.00	2.74	6.35	80.00	1.30	0.4611
LNEI		0.00	8.00	8.00	38.00	303.00	17.00	41.00	8.00	55.0	262.00	2.00	5.74	41.00	3.40	0.4629
WA	46.41	1.35	11.96	6.93	56.48	237.52	30.94	39.91	9.54	53.53	217.63	2.59	6.13	52.72	2.29	
Mean	46.09	1.33	11.67	6.89	56.78	238.06	30.72	39.69	9.44	53.39	217.72	2.58	6.13	53.06	2.29	
Mean		3.1		15.3	42.5	257.7	39.2	37.8	8.5	60.8	232.2	2.9	6.39	40.8	2.2	
Max		7.0		40.0	70.0	664.8	85.0	89.0	15.4	158.3	495.7	7.9	6.95	56.3	5.2	
Min		<2.5		<2.5	17.5	123.9	21.0	4.0	2.6	25.0	130.4	<1.4	5.75	27.1	<0.1	

Table 48: Diatom and Cladocera-based modern analogues for Blue Lough

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	TOC	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
CADH	6.50	3.50	16.00	19.50	49.00	354.20	52.00	-2.00	9.00	64.00	288.0	0.00	5.14	55.50	5.35	0.4534
WA	6.50	3.50	16.00	19.50	49.00	354.20	52.00	-2.00	9.00	64.00	288.0	0.00	5.14	55.50	5.35	
Mean	6.50	3.50	16.00	19.50	49.00	354.20	52.00	-2.00	9.00	64.00	288.0	0.00	5.14	55.50	5.35	
Mean		286.9		377.2	40.0	275.8	55.9	-22.8	12.8	60.0	257.0	28.6	4.69	94.8	3.5	
Max		470.0		520.0	98.0	400.0	73.0	-4.0	25.4	91.7	39.6	72.9	5.11	118.8	6.8	
Min		72.0		280.0	16.5	152.1	36.0	-33.0	7.7	33.3	121.7	10.0	4.51	35.4	1.4	

Table 49: Diatom and Cladocera-based modern analogues for Llyn Llagi

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	Toc	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
ARR	55.17	2.00	24.00	8.67	43.33	219.17	39.17	49.06	13.33	59.17	216.17	3.00	6.30	38.17	2.17	0.4547
WA	55.17	2.00	24.00	8.67	43.33	219.17	39.17	49.06	13.33	59.17	216.17	3.00	6.30	38.17	2.17	
Mean	55.17	2.00	24.00	8.67	43.33	219.17	39.17	49.06	13.33	59.17	216.17	3.00	6.30	38.17	2.17	
Mean		39.7		74.1	52.5	193.8	31.2	5.6	6.2	46.7	168.7	10.0	5.34	61.0	2.4	
Max		159.0		193.0	94.0	377.5	58.0	33.4	19.2	75.0	291.3	38.6	6.30	81.3	5.50	
Min		<2.5		5.0	31.0	98.6	13.0	-8.0	2.6	25.0	100.0	2.1	4.78	39.6	<0.10	

Table 50: Diatom and Cladocera-based modern analogues for Llyn cwm Mynach

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	TOC	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
LOSG	18.50	3.50	30.00	33.50	44.50	146.50	25.00	3.00	5.00	30.50	122.00	0.00	5.53	21.00	6.60	0..3328
DUBH		2.25	9.25	11.50	59.00	741.25	96.50	9.00	13.25	112.50	585.75	0.25	5.62	75.00	3.25	0.3824
TINK	31.33	4.67	18.00	19.00	79.25	141.75	29.25	24.17	8.00	38.75	121.00	5.00	5.72	64.25	3.60	0.3880
FEOI		1.25	12.25	13.50	58.25	498.75	70.00	35.75	11.50	89.25	400.00	0.00	6.12	50.25	4.55	0.3904
TEAN	20.00			0.00	110.00	1160.00	16.30	12.04	23.00	255.00	981.00	7.00	5.70	233.00		0.4038
NEUN		1.00	5.75	6.75	30.75	383.75	53.50	7.75	8.25	61.50	312.50	0.00	5.68	40.25	3.08	0.4114
ARR	55.17	2.00	24.00	8.67	43.33	219.17	39.17	49.06	13.33	59.17	216.17	3.00	6.30	38.17	2.17	0.4281
CLAI		2.00	11.00	13.00	37.00	245.50	33.00	34.00	7.50	34.00	193.50	0.00	6.13	31.00	2.20	0.4375
ACH	12.25	3.00	14.00	6.60	41.80	198.40	36.40	4.33	7.00	32.60	210.40	1.25	5.24	36.60	2.75	0.4494
LACH	32.00	17.00	38.00	55.00	110.00	155.00	33.00	32.00	7.00	50.00	139.00	1.00	5.92	71.00	5.80	0.4699
WA	27.81	3.89	16.21	16.74	60.94	392.50	43.45	20.50	10.37	76.71	330.15	1.74	5.79	65.78	3.85	
Mean	28.21	4.07	18.03	16.75	61.39	389.01	43.21	21.11	10.38	76.33	328.13	1.75	5.80	66.05	3.78	
Mean		65.2		117.3	70.0	304.2	46.0	4.6	5.6	63.3	268.3	10.0	5.37	86.0	2.6	
Max		291.0		378.0	128.0	518.3	72.0	34.4	9.7	100.0	404.3	30.7	6.30	154.2	10.7	
Min		<2.5		5.0	21.5	143.7	24.0	-21.0	2.6	33.3	173.9	2.1	4.70	58.3	<0.1	

Table 51: Diatom and Cladocera-based modern analogues for Lochnagar

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	TOC	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
LOCH		0.00	5.00	5.00	22.50	260.00	34.50	15.50	10.00	39.00	204.50	0.00	5.99	35.50	1.05	0.4474
CLYD	62.24	7.33	8.33	16.25	55.90	133.98	29.25	33.93	5.31	48.00	123.90		6.15	65.75	0.48	0.4772
WA	62.44	3.55	6.61	10.44	38.66	199.02	31.96	24.42	7.73	43.35	165.50	0.00	6.07	50.14	0.77	
Mean	62.44	3.67	6.67	10.63	39.20	196.99	31.88	24.72	7.66	43.50	164.20	0.00	6.07	50.63	0.77	
Mean		25.5		41.8	29.0	89.3	21.8	0.6	7.4	33.3	93.9	15.7	5.33	57.7	1.1	
Max		137.0		147.0	50.0	166.2	35.0	12.0	12.8	58.3	173.9	30.7	5.81	85.4	3.4	
Min		<2.5		4.0	21.5	50.7	4.0	-10.0	2.6	25.0	69.6	<1.4	4.95	45.8	0.2	

Table 52: Diatom and Cladocera-based modern analogues for Round Loch of Glenhead

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	TOC	DC
	µ eq l^{-1}	µ g l^{-1}	µ g l^{-1}	µ g l^{-1}	µ eq l^{-1}	µ eq l^{-1}	µ S cm^{-1}	µ eq l^{-1}	µ eq l^{-1}	µ eq l^{-1}	µ eq l^{-1}	µ eq l^{-1}		µ eq l^{-1}	Mg l^{-1}	
LNEI		0.00	8.00	8.00	38.00	303.00	17.00	41.00	8.00	55.00	262.00	2.00	5.74	41.00	3.40	0.3873
DUBH		2.25	9.25	11.50	59.00	741.25	96.50	9.00	13.25	112.50	585.75	0.25	5.62	75.00	3.25	0.3879
CLAI		2.00	11.00	13.00	37.00	245.50	33.00	34.00	7.50	34.00	193.50	0.00	6.13	31.00	2.20	0.4074
LACH	32.00	17.00	38.00	55.00	110.00	155.00	33.00	32.00	7.00	50.00	139.00	1.00	5.92	71.00	5.80	0.4107
DOI	22.17	7.00	47.67	27.33	54.00	230.83	43.83	15.43	12.50	60.50	238.67	6.40	5.77	58.83	2.40	0.4409
LAI	20.75	2.67	21.67	18.25	41.25	130.00	24.75	121.81	6.25	31.00	132.00	2.75	5.79	45.50	3.10	0.4553
HIR2	14.00	7.30	18.00	25.00	74.00	182.00	35.00	5.00	6.00	65.00	171.00	63.00	5.57	77.00	3.10	0.4690
CHAM		9.40	12.00	21.40	49.40	530.40	72.80	6.20	10.60	87.20	412.40	1.00	5.60	59.60	2.95	0.4738
ARR	55.17	2.00	24.00	8.67	43.33	219.17	39.17	49.06	13.33	59.17	216.17	3.00	6.30	38.17	2.17	0.4743
WA	24.15	5.44	20.80	20.85	56.33	309.22	44.03	23.04	9.38	61.82	264.65	8.22	5.83	55.21	3.18	
Mean	28.82	5.51	21.07	20.91	56.22	304.13	43.89	22.72	9.38	61.60	261.17	8.82	5.83	55.23	3.15	
Mean		60.2		95.1	33.0	195.2	36.7	-12.2	8.2	45.8	174.3	7.1	4.90	67.5	3.0	
Max		110.0		146.0	42.0	298.6	49.0	6.0	12.8	66.7	247.8	24.3	5.21	114.6	5.0	
Min		9.0		55.0	25.0	121.1	28.0	-22.0	2.6	33.3	130.4	<1.4	4.72	45.8	1.6	

Table 53: Diatom and Cladocera-based modern analogues for Scoat Tarn

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	TOC	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
EDNO	25.31	54.5	23.00	78.33	38.07	150.43	28.67	0.00	5.20	28.87	134.43	8.90	5.15	59.7	1.43	0.3981
WA	25.31	54.5	23.00	78.33	38.07	150.43	28.67	0.00	5.20	28.87	134.43	8.90	5.15	59.7	1.43	
Mean	25.31	54.5	23.00	78.33	38.07	150.43	28.67	0.00	5.20	28.87	134.43	8.90	5.15	59.7	1.43	
Mean		108.9		121.4	32.5	185.6	35.1	-8.4	7.9	48.3	163.9	21.4	4.99	60.8	0.9	
Max		293.8		300.0	48.0	326.8	49.0	6.0	15.4	75.0	265.2	47.9	5.23	72.9	2.7	
Min		2.5		12.2	23.0	118.3	24.0	-26.0	2.6	33.3	126.1	5.7	4.57	35.4	<0.1	

Table 54: Diatom and Cladocera-based modern analogues for Loch Tinker

Lake	Alk1	Allab	Almon	Altot	Ca	Cl	Cond	EqAlk	K	Mg	Na	NO_3	pH	SO_4	TOC	DC
	μ eq l^{-1}	μ g l^{-1}	μ g l^{-1}	μ g l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ S cm^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}	μ eq l^{-1}		μ eq l^{-1}	Mg l^{-1}	
TINK	31.33	4.67	18.00	19.00	79.25	141.75	29.25	24.17	8.00	38.75	121.00	5.00	5.72	64.25	3.60	0.3683
ACH	12.25	3.00	14.00	6.60	41.80	198.40	36.0	4.33	7.00	32.60	210.40	1.25	5.24	36.60	2.75	0.3849
LAI	20.75	2.67	21.67	18.25	41.25	130.00	24.75	12.81	6.25	31.00	132.00	2.75	5.79	45.50	3.10	0.3858
LOSG	18.50	3.50	30.00	33.50	44.50	146.50	25.00	3.00	5.00	30.50	122.00	0.00	5.53	21.00	6.60	0.3892
FHIO		6.25	24.75	31.00	29.50	429.75	61.00	-9.25	10.25	73.50	334.50	0.00	4.97	49.00	3.35	0.3898
NEUN		1.00	5.75	6.75	30.75	383.75	53.50	7.75	8.25	61.50	312.50	0.00	5.68	40.25	3.08	0.4170
GRUA		3.75	17.25	21.00	18.75	282.50	40.25	-3.75	8.25	43.00	235.75	0.25	5.16	34.75	2.85	0.4605
WA	20.83	3.58	18.92	19.50	41.75	241.34	38.31	5.93	7.55	44.18	207.29	1.39	5.45	41.97	3.64	
Mean	20.71	3.55	18.77	19.44	40.83	244.66	38.59	5.58	7.57	44.41	209.74	1.32	5.44	41.62	3.62	
Mean		3.3		20.3	85.0	162.8	31.4	37.6	8.5	48.3	142.6	2.9	6.13	55.6	4.7	
Max		14.0		45.0	127.0	439.4	62.0	96.0	17.9	91.7	321.7	14.3	6.56	110.4	8.1	
Min		<2.5		5.0	35.0	70.4	21.0	-2.0	2.6	33.3	78.3	<1.4	5.42	37.5	1.9	

From Table 46 we can see that the combined Cladocera and diatom training set contains close modern analogues for only some of the UKAWMN lakes. Pre-acidification conditions in Burnmoor Tarn and Loch Chon are characterised by a diatom flora and cladoceran fauna that appear to be quite distinct from the flora and fauna of the samples in the training set and as such no close analogues for these two lakes can be identified.

A number of UKAWMN lakes have a diatom and cladoceran community that is well represented in the training set. Llyn cwm Mynach, the Round Loch of Glenhead and Loch Tinker each have 10, 9 and 7 close modern analogues in the training set respectively for the sub fossil assemblages found in their pre-acidification sediments.

Loch Coire nan Arr is represented by 3 close modern analogues and Lochnagar by just 2, whilst the remaining lakes, Blue Lough, Llyn Llagi and Scoat Tarn, each have only a single close modern analogue in the 83-lake training set.

The results presented in Table 46 show a number of interesting features. In general, there are fewer close modern analogues for a particular UKAWMN lake using the combined diatom and cladoceran training set than with either the 163- or 83-lake diatom training sets. The main exceptions to this pattern are for those UKAWMN lakes that have few or no close modern analogues in the diatom training sets (Blue Lough, Lochnagar and Scoat Tarn).

For the two lakes with no close modern analogues in the combined training set, Table 46 shows that the lack of analogues is driven mainly by the diatom flora in Burnmoor Tarn and by the cladoceran fauna in Loch Chon. The pre-acidification sediments of Burnmoor Tarn are abundant in *Achnanthes minutissima* (20%+) and whilst this diatom occurs in many of the samples in the respective diatom training sets, it is very uncommon that this species is found at such high abundances. A further characteristic of the pre-acidification sediments in Burnmoor Tarn is the planktonic diatom component. The dominant diatom taxon in the pre-acidification sample from Burnmoor Tarn is *Cyclotella comensis*, where as across the training set as a whole, *Cyclotella* [*kuetzingiana* agg.] is generally more abundant in those lakes with a planktonic diatom community.

This is perhaps the result of taxonomic harmonisation when the training sets were compiled from the existing diatom data. The majority of samples in the 163- and 83-lake training sets are taken from the SWAP training set where few centric diatom taxa were recorded to species level. In the *C. kuetzingiana* species group in particular the taxonomy of the SWAP training set does not distinguish between the nominate and its varieties, lumping the varieties in the *C. kuetzingiana* group into a *Cyclotella* [*kuetzingiana* agg.] group (Stevenson *et al.* 1991). Those samples counted after then period of SWAP benefited from a more well-defined taxonomy for these centric taxa and differentiated between a wider range of centric species. When the UKAWDDS was compiled from the UK SWAP samples and those samples from UK lakes counted subsequently it was necessary to harmonise these new samples to SWAP taxonomy; the direct result of which is that *Cyclotella* [*keutzingiana* agg.] is the dominant centric taxon in the two diatom training sets, either by virtue of *C. kuetzingiana* being found in those samples or because of the harmonisation process.

A further reason that might explain the lack of close modern analogues for Burnmoor Tarn is that chemically the lake is quite different from the majority of lakes in the training set. Burnmoor Tarn is a naturally alkaline system with a mean pH of 6.48 (1988-1998), and whilst the site has had a diatom-inferred pH of *ca.* 6.5 throughout the 20th Century the lake is subject to strong acid episodes (min measured pH of 4.38 over the period 1988-1998). The lake is also better buffered than the majority of samples in the training set (mean lake-water calcium concentration = 89 μ eq l^{-1}). This type of lake is poorly represented in the training set and from the few lakes that are hydrochemically similar to Burnmoor Tarn it is not surprising that close modern diatom analogues cannot be found.

The lack of close modern analogues for the pre-acidification sample in Loch Chon is driven mainly by the cladoceran community composition of the sample. This sample is quite different from the majority of cladoceran samples in the 83-lake training set for two reasons. The first reason is that *B. coregoni* is not present in the sample at all. Of the 83-lakes in the training set, only eight of these are similarly devoid of this planktonic cladoceran. The second reason is that the sample from Loch Chon is very rich in chydorid taxa (15 species) and includes a number of large bodied, macrophyte associated taxa such as *G. testudinaria*, *E. lamellatus* and *S. crystallina.* In those samples from the training set where *B. longispina* is the dominant planktonic cladoceran, the majority of samples are not as chydorid-rich as this sample from Loch Chon.

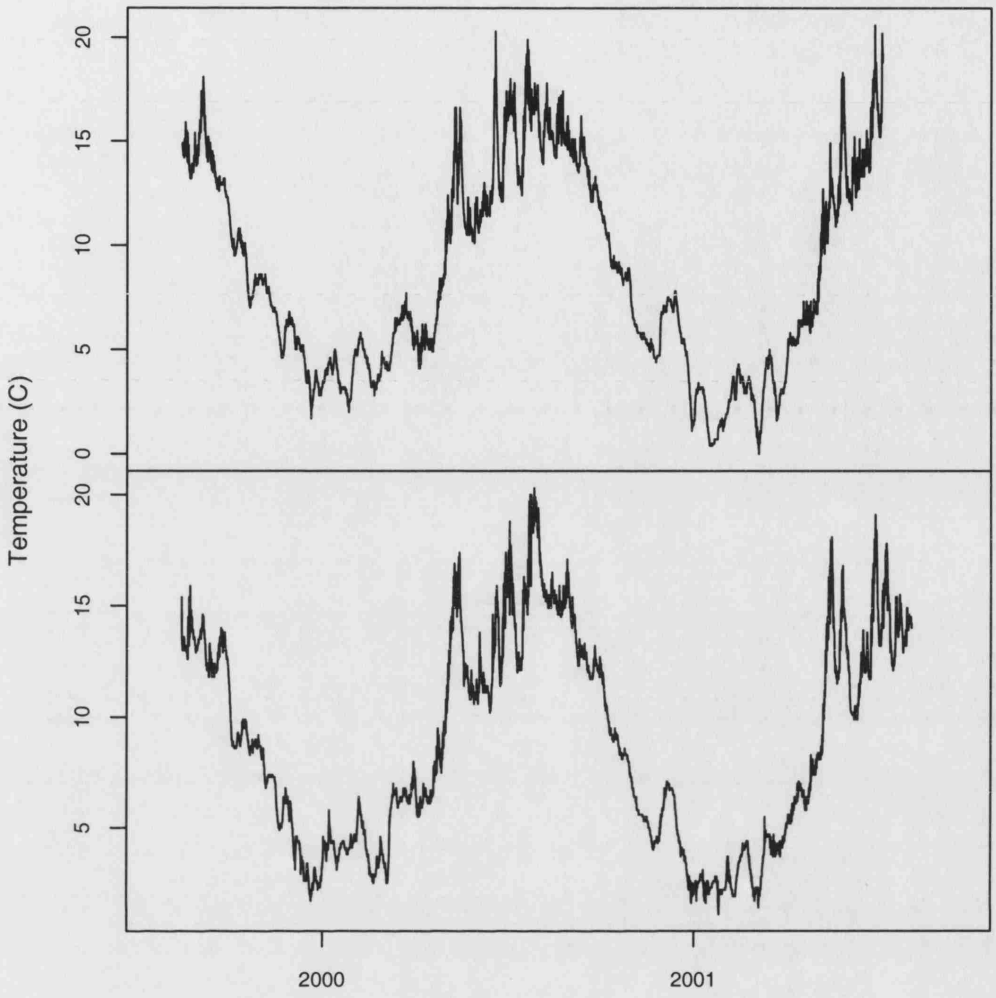


Figure 87: Plot showing time series data from Llyn Llagi (top) and Loch Coire nan Arr (bottom) of 2-hourly lake surface temperature recordings from late 1999 to mid 2001. The data were recorded by a data logger attached to a sediment trap located in the deepest area of each lake.

Unlike Burnmoor Tarn, Loch Chon is typical of the upland, acid-sensitive lakes found in the UK. It is mildly acidic (ten year mean pH of 5.60) with low acid neutralising capacity (ten year mean alkalinity of 10 μ eq l^{-1}) and moderate sea-salt inputs. It is not immediately clear why there are no close modern analogues in the training set, though this may just be due to insufficient coverage of the environmental space by the training set.

The question that arises therefore is *how reliable is the analogue?* The first step is to compare the hydrochemistry for the analogue and the present day conditions in the acidified lake in question. The hydrochemistry for the modern analogues should be similar to that measured at the present day in terms of the major ions and conductivity. Secondly, aluminium concentrations should be lower in the modern analogue lakes than those

measured in the acidified lake as aluminium becomes more available as pH decreases. Thirdly, the pH and alkalinity of the lake water should be higher in the modern analogue than in the acidified lake. Furthermore, the lake water calcium concentration should be approximately the same in the two lakes but preferably lower in the modern analogues because calcium concentrations in lake water can be increased as a lake acidifies due to increased base cation leaching from soils. A further possibility is to compare the species present in the sub-fossil assemblages of the pre-impact sample in the acidified lake and the relevant modern analogue, and compare the diversity of the two samples. Finally one can use existing secondary sources of information, where they exist, to try to validate the modern analogues. However, given the remoteness of these upland lakes historical records of the species present in the early 1800's, prior to the onset of acidification in the UK, are generally lacking.

Table 47 to Table 54 list the modern analogues identified for each of the UKAWMN lakes for which a close modern analogue can be found in the diatom and cladoceran training set. The hydrochemical properties of each of these analogue lakes are presented alongside summary data of the measured hydrochemistry for each of the UKAWMN lakes.

5.4.2.1 Llyn Llagi

In the case of Llyn Llagi (Table 49) the hydrochemistry of the single close modern analogue compares favourably, in terms of the major ions and the conductivity, with the measured hydrochemistry over the last 10 years in Llyn Llagi. As well as this close agreement, the modern analogue, Loch Coire nan Arr, suggests that the hydrochemistry of Llyn Llagi was much less acidic in the past (inferred pH = 6.3, mean measured pH = 5.35 between 1988 and1998) with much lower labile aluminium concentrations coupled with a significantly large lake-water alkalinity concentration. The lake-water calcium concentration in Llyn Llagi is also larger than in Coire nan Arr, its closest modern analogue. These data suggest that Loch Coire nan Arr is hydrochemically similar to that of Llyn Llagi prior to Llyn Llagi becoming acidified and are consistent with the kind of hydrochemistry one would expect to have been characteristic of the lake prior to the onset of disturbance.

Llyn Llagi is situated at an altitude of 380 metres above sea level (m i.e.), however, which is significantly higher than Loch Coire nan Arr (125 m i.e.). This difference in altitude might

indicate that whilst the Loch Coire nan Arr is a close diatom and cladoceran analogue for pre-acidification conditions in Llyn Llagi, climatic differences between the two lakes that relate to the difference in altitude might have significant implications for the wider biological groups supported by the two lakes. Any temperature difference related to altitude may be offset by the difference in latitude between the two lakes, with Loch Coire nan Arr located in the far North West of Scotland compared to the North Wales location of Llyn Llagi.

This can be assessed climatic data collected from the two lakes. Figure 87 shows the measured, 2-hourly surface temperature data from late 1999 to mid 2001 in the two lakes. The two temperature records are highly correlated and suggest that despite a difference of over 200 metres in altitude between the lakes, they both have a similar temperature regime.

The species diversity, richness and evenness indices for Loch Coire nan Arr and the core bottom sample from Llyn Llagi also show a similar level of agreement (Table 55). The two samples contain a similar level of species diversity (3.215 in Loch Coire nan Arr and 3.267 in Llyn Llagi), with the core bottom sample from Llyn Llagi being slightly less species rich than Coire nan Arr (62 and 67 species present respectively). This difference in species richness is likely due to the presence of a greater number of taxa that are present at low abundances in Loch Coire nan Arr, indicated by the lower value for Pylon's evenness in the sample. Pylon's evenness measures how evenly distributed the species abundances in a single sample are, with higher values for indicating a more-even distribution of taxa in the sample.

The results of the analogue matching for Llyn Llagi have identified a single close modern analogue lake; Loch Coire nan Arr. The results presented in Table 49, Tbale 55 and Figure 87 suggest that, as well as being biologically similar to Llyn Llagi before the onset of acidification, Loch Coire nan Arr is similar to Llyn Llagi in terms of the lake-water major ion composition, thermodynamic regime and species diversity.

5.4.2.2 Blue Lough

Like Llyn Llagi, only a single was identified for each of Blue Lough in Northern Ireland and Scoat Tarn in the English Lake District. Loch Dubh Cadhafuaraich in northern

Scotland is the closest modern analogue for Blue Lough and Llyn Edno in Snowdonia, North Wales, is the closest analogue for Scoat Tarn.

The main contributors to the similarity between Loch Dubh Cadhafuaraich and the pre-acidification sample from Blue Lough are the high abundances of *Eunotia incisa*, *Frustulia rhomboides* var. *saxonica*, and *Cymbella perpusilla* present in both samples. Furthermore, despite both lakes being naturally acidic systems *Tabellaria quadriseptata* is absent from Loch Dubh Cadhafuaraich and present at slightly greater than 2% in the pre-acidification sample from Blue Lough. The sample from Loch Dubh Cadhafuaraich is much more diverse than that from Blue Lough, containing a more even distribution of taxa (Pylon's evenness = 0.815 in Loch Dubh Cadhafuaraich and 0.793 in Blue Lough) and a greater number of species (54 species in Loch Dubh Cadhafuaraich compared to 39 in Blue Lough). The hydrochemistry of Loch Dubh Cadhafuaraich (Table 48) is quite similar to that record between 1988 and 1998 in Blue Loch, with the exception that Loch Dubh Cadhafuaraich has a much lower labile aluminium concentration and a higher alkalinity and pH than that recorded in Blue Lough. The two lakes are quite similar in terms of the concentrations in major ions in the hydrochemistry, including the lake-water calcium concentration. The analogue-inferred pH for pre-acidification conditions in Blue Lough is 5.14. Compared with the diatom-inferred pH reconstruction for this sample, pH *ca.* 4.9 (Patrick, Monteith, and Jenkins 1995), this value is a little high though not significantly so.

Loch Dubh Cadhafuaraich is located in an area of the UK that receives low levels of acid deposition, and whilst the differences in species diversity between the pre-disturbance sample from Blue Lough and the surface sample from Loch Dubh Cadhafuaraich might indicate caution in accepting Loch Dubh Cadhafuaraich as a valid analogue for Blue Lough, this low level of disturbance, coupled with the hydrochemical inferences and the similarities in diatom and cladoceran species composition, would suggest that Loch Dubh Cadhafuaraich is a suitable modern analogue for Blue Lough.

5.4.2.3 Scoat Tarn

The similarity between the Scoat Tarn pre-disturbance sediment sample and its modern analogue, Llyn Edno, is driven mainly by the abundance of *C.* [*keutzingiana* agg.] and of a number of Cladocera taxa. The surface sample from Llyn Edno is dominated by *C.*

[*keutzingiana* agg.], which is present in the sample at over 65% abundance. This species group is also the dominant diatom taxon in the pre-acidification sample from Scoat Tarn, though it is only present at *ca.* 25% abundance in this sample. The proportions of *B. coregoni* in the two samples are very similar, with this species being present at *ca.* 25% in both Llyn Edno and the pre-acidification sample from Scoat Tarn.

Both of the samples also contain similar proportions of three taxa commonly found in alpine or montane lakes; *A. harpae*, *A. nana* and *A. excisa*. Llyn Edno lies at an altitude of 500 m i.e. whilst Scoat Tarn lies at an altitude of 602 m i.e., and as such are high altitude lakes in the context of UK upland fresh water systems. The two samples are not similarly diverse in their diatom and cladoceran communities, however, with the sediment sample from Llyn Edno containing a greater number of taxa than that from Scoat Tarn (Table 55).

Table 55: Table showing the Shannon diversity index, the species richness and Pylon's evenness index for the close modern analogues of the pre-disturbance sediments in the seven UKAWMN lakes for which a close modern analogue exists in the diatom and Cladocera training set. The three measures are shown for the core sample taken from each respective UKAWMN lake so that comparison can made between the modern analogues and each impacted lake.

Loch Coire nan Arr

Analogue	Diversity	Richness	Evenness
ARR	3.215	67	0.765
CWEL	3.037	62	0.736
LNEI	3.367	72	0.787
Core Bottom	2.965	64	0.713

Llyn Llagi

Analogue	Diversity	Richness	Evenness
ARR	3.215	67	0.765
Core Bottom	3.267	60	0.798

Lochnagar

Analogue	Diversity	Richness	Evenness
LOCH	3.554	75	0.823
CLYD	3.075	48	0.794
Core Bottom	3.331	59	0.817

Roundloch of Glen Head

Analogue	Diversity	Richness	Evenness
LNEI	3.367	72	0.787
DUBH	3.417	66	0.816
CLAI	3.341	61	0.813
LACH	3.529	71	0.828
DOI	3.473	68	0.823
LAI	3.078	58	0.758
HIR2	3.598	71	0.844
CHAM	3.393	61	0.825
ARR	3.215	67	0.765
Core Bottom	3.626	71	0.851

Blue Lough

Analogue	Diversity	Richness	Evenness
CADH	3.250	54	0.815
Core Bottom	2.905	39	0.793

Llyn Cwm Mynach

Analogue	Diversity	Richness	Evenness
LOSG	3.522	66	0.841
DUBH	3.417	66	0.816
TINK	3.092	54	0.775
FEOI	3.301	63	0.797
TEAN	3.265	57	0.808
NEUN	3.232	55	0.807
ARR	3.215	67	0.765
CLAI	3.341	61	0.813
ACH	3.272	51	0.832
LACH	3.529	71	0.828
Core Bottom	3.518	65	0.848

Scoat Tarn

Analogue	Diversity	Richness	Evenness
EDNO	2.818	67	0.669
Core Bottom	3.155	59	0.774

Loch Tinker

Analogue	Diversity	Richness	Evenness
TINK	3.092	54	0.775
ACH	3.272	51	0.832
LAI	3.078	58	0.758
LOSG	3.522	66	0.841
FHIO	3.208	53	0.808
NEUN	3.232	55	0.807
GRUA	3.469	58	0.854
Core Bottom	2.835	51	0.721

Hydrochemically, the same pattern described for Llyn Llagi and its closest modern analogue exists for Scoat Tarn and Llyn Edno. Table 53 shows the hydrochemistry data for Llyn Edno as well as the 10-year data for Scoat Tarn. Inferring the pre-acidification hydrochemistry of Scoat Tarn from that of Llyn Edno would suggest a pre-disturbance pH of 5.15. The diatom-inferred pH from this sample is *ca.* 5.75 (Patrick *et al.* 1995), driven mainly the large proportions of *C.* [*keutzingiana* agg.] and *Achnanthes minutissima* in the sample which both have pH optima of 6.3 (Stevenson *et al.* 1991). So there is a mismatch between the analogue-inferred pH and the diatom-inferred pH.

Unlike the modern analogues for Llyn Llagi and Blue Lough, less confidence can be placed in the modern analogue for Scoat Tarn. Despite the similarities in the planktonic diatom and cladoceran species, large discrepancies between the two samples exist. *B. longispina* contributes a much larger proportion of the planktonic cladoceran fauna of Scoat Tarn (36% abundant) than Llyn Edno (8% abundant). Indeed, in Scoat Tarn, *B. longispina* is the dominant cladoceran where as in Llyn Edno, *B. coregoni* is dominant. As has already been shown in Chapter 4, at the large scale, lakes where *B. longispina* is the more dominant taxon tend to have a different cladoceran fauna to those lakes where *B. coregoni* is dominant.

In Scoat Tarn, *A. minutissima* and *Achnanthes marginulata* are both quite abundant in the core bottom sample. These two taxa are found in Llyn Edno but at much lower abundances than in Scoat Tarn. Indeed, this combination of the high abundances of *A. minutissima*, *A. marginulata* and *C.* [*keutzingiana* agg.] is a particularly rare occurrence across the training set as a whole. Using the 163-lake and the 83-lake diatom training sets, no close modern analogues could be identified (Table 46),.whilst the cladoceran fauna of Scoat Tarn is very similar to 9 other lakes in the 82-lake training set (Table 46). This suggests that it is the diatom flora of the pre-acidification sample from Scoat Tarn that is very different from other upland lakes in the UK. It remains to be seen whether this difference is a perceived result of the limited population of lakes in the UK found in the training set or that the difference is due to a unique combination of environmental and physiological factors that might exist at Scoat Tarn.

These factors suggest that despite containing similar species and that on the whole, the diatom and cladoceran "*community*" of the two samples is considered similar (squared chord dissimilarity = 0.3981), there are significant differences in the proportions of some of the

taxa in the two samples. Coupled with the fact that Llyn Edno is much more acidic than the diatom composition of the pre-acidification sample of Scoat Tarn would suggest, it is unlikely that Llyn Edno would be similar to a pre-disturbance Scoat Tarn with regard to other biological groups. This assumption cannot currently be answered categorically as little is known about wider biological groups (e.g. aquatic macrophytes and invertebrates), in either Scoat Tarn prior to the onset of acidification or in Llyn Edno at the present day. Work currently being undertaken for the Department of Environment, Food and Rural Affairs aims to investigate similarities between these wider biological groups and those that existed prior to acidification by extending the range of proxy records analysed from lake sediments in both the acidified lakes and those identified as modern analogues.

5.4.2.4 Loch Coire nan Arr

Three modern analogues were identified from the 83-lake diatom and cladoceran training set for the pre-acidification sample from Loch Coire nan Arr. Interestingly, Loch Coire nan Arr is one of two UKAWMN lakes whose surface sediment samples are included in the diatom and cladoceran training set (the other being Loch Tinker) and this surface sample from Loch Coire nan Arr is identified as being the closest modern analogue for the pre-acidification sample in the same loch. Llyn Cwellyn in North Wales and Loch nan Eion in North West Scotland are the other two modern analogues identified. The surface sample from Loch Coire nan Arr is considerably more similar (squared chord dissimilarity = 0.4297) to the pre-acidification sample than either of the other two analogues (squared chord dissimilarity = 0.4611 and 0.4629 for Llyn Cwellyn and Loch nan Eion respectively), though all three samples are comfortably within the 5% cut-off value used.

This result is to be expected as Loch Coire nan Arr is a minimally impacted lake with respect to acid deposition. Patrick *et al.* (1995) and Monteith and Evans (2000) reinterpreted the palaeoecological diatom record from Loch Coire nan Arr and suggest that, contrary to earlier evidence (Flower *et al.* 1993), Loch Coire nan Arr has acidified, though only slightly. This acidification appears more likely to be a natural response to changes in the prevailing climate in North West Scotland over the last 70 years or so (Monteith, 2002 *pers. comm.*) than the effects of acid deposition *per se*.

In a recent paper by Rose and Rippey (2002) the historical record of polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), trace metals and fly-ash particle deposition at Loch Coire nan Arr was investigated. The sediment record of fly-ash particles (SCPs), which are only produced from the burning of fossil fuels at high temperatures, does indicate that Loch Coire nan Arr is not a pristine ecosystem. The levels of SCPs found in the sediment record are low, however, compared to other areas in the UK, peaking at *ca.* 12 500 SCPs per gram dry weight of sediment in the late 1970s. Other trace metal records indicate little change in levels throughout the recent (1800 to present day) history of the loch, and peak concentrations of PAHs in the sediment record were equivalent to natural, background levels recorded at sites in the English Lake District. Rose and Rippey (2002) conclude that Loch Coire nan Arr is thought to be one of the least polluted lakes in the UK and that the levels of PCBs recorded in the sediment of the loch are amongst the lowest found in Europe.

Loch Coire nan Arr is also a close modern analogue for two other UKAWMN lakes; Llyn Llagi (See above) and Llyn cwm Mynach. It is clear from the accumulated data on this upland loch that it is a minimally impacted site with respect to acid deposition. Routine monitoring work at the loch from 1988 to 1998 as part of the UKAWMN can provide quite detailed records of the kinds of biology likely to have occurred in pristine acid sensitive surface waters in the UK. It is unfortunate, however, that this loch has been disturbed through the introduction of a permanent dam and sluice at one end of the loch as a means of providing a controlled source of water for downstream fish farming activities (Monteith and Evans 2000). In recent years this impact has resulted in the raising of the lake water level by *ca.* 0.5 metres leading to the loss of the loch's emergent macrophyte community and the flooding of heather moorland around the margins of the loch. As such, Loch Coire nan Arr can no longer be considered as a reliable analogue for similar acidified sites in the UK. The monitoring data collected at the site prior to the engineering work will, however, be invaluable in determining the likely biological community types found in acidified lakes for which Loch Coire nan Arr is a close modern analogue.

The similarity between the fossil assemblage from Loch Coire nan Arr and the surface samples from Llyn Cwellyn and Loch nan Eion is derived from similarities in both the diatom and cladoceran communities. All three sediment samples are dominated by planktonic Cladocera. In Loch Coire nan Arr, *B. longispina* is the dominant cladoceran,

whilst in both Llyn Cwellyn and Loch nan Eion *B coregoni* is dominant. Both taxa are present at high abundances in all three samples, as is *C. piger*, which is present at over 5% of the cladoceran community. *A. minutissima* var. *minutissima* is the most abundant diatom taxon in Loch Coire nan Arr and Llyn Cwellyn and is present at *ca.* 7% in Loch nan Eion. All three samples are characterised by high proportions of this taxon, as well as *Brachysira vitrea* and *Tabellaria flocculosa* var. *flocculosa.* All samples have few planktonic diatom taxa present and acidobiontic taxa, such as *T. quadriseptata* and *Eunotia incisa* are either not present or found at only very low abundances.

Loch Coire nan Arr and Llyn Cwellyn are similarly rich in species (67 diatom taxa and 16 cladoceran species in Loch Coire nan Arr compared to 62 and 17 in Llyn Cwellyn), with Loch nan Eion being slightly more species rich (72 diatom and 21 cladoceran species). These figures are derived from the full species list present in the samples but compare favourably to the results of the diversity measures illustrated in Table 55, which are derived from only those taxa present at 2% or more in a single sample in the training set.

The hydrochemical data from the three close analogues for Loch Coire nan Arr are shown in Table 47 above. All three modern analogues are similar hydrochemically, though Loch nan Eion is slightly more acid than the other two analogues with a mean measured pH of 5.74, which compares with the diatom-inferred pH of *ca.* 6.1 for the core bottom sample. The pH of Loch Coire nan Arr (pH = 6.3) and Llyn Cwellyn (pH = 6.35) shows a better fit to the diatom inferred pH. All three lakes are well buffered with high alkalinities and relatively high lake water calcium concentrations though these are well within the range of values measured in Loch Coire nan Arr over the period 1988 to 1998.

The data and findings presented above all point to the fact that Loch Coire nan Arr, whilst not being pristine, can be classed as a minimally impacted upland fresh water. The diatom and cladoceran community present in the loch today is numerically very similar to the core-bottom sample and is the closest match in the training set for the core bottom. All three of the close modern analogues are floristically and faunistically similar, with comparable levels of species diversity, richness and evenness which are analogous to these measures for the core bottom sample.

Based on the biological monitoring in Loch Coire nan Arr as part of the ongoing work of the UKAWMN it is likely that the Loch Coire nan Arr of the early nineteenth century was characterised by an aquatic macrophyte flora typical of non-acidified, oligotrophic upland lakes containing the acid-sensitive charophyte *Nitella flexilis* and *Carex rostrata*, a species which has subsequently been lost from the loch since the water level was raised in 1992. *Sphagnum auriculatum* would probably occur though only in low abundances. The invertebrate fauna of the loch would be relatively diverse and would have contained a number of taxa intolerant of very acid conditions, such as *Lymnaea peregra*, *Pisidium* spp. and *Baetis* spp. Trout are also likely to have been present.

Inferring the hydrochemistry of past periods in lake systems from the hydrochemistry found today in close modern analogues lakes is a simple form of calibration. The methodology is discussed more fully in Chapter 2. Table 47 shows the mean hydrochemistry of the three modern analogues as well as the dissimilarity-weighted mean hydrochemistry. These values can be taken as being the inferred hydrochemistry for *ca.* 1800 in Loch Coire nan Arr. The dissimilarity-weighted mean values give greater weight to the most similar sample by taking the inverse of the squared chord dissimilarity as the weights in the calculations. Both methods reconstruct the *ca.* 1800 pH in Loch Coire nan Arr as being 6.13. This value compares favourably with the weighted average diatom-inferred pH of 6.1 (Patrick *et al.* 1995).

The reconstructed values for the other hydrochemical parameters shown in Table 47 can also be taken as being valid inferences for the past hydrochemistry of Loch Coire nan Arr, however the methodology for providing sound estimates of the error inherent in these predictions is not as advanced as for the family of weighted average and weighted average partial least squares calibration techniques. Ter Braak (1995) suggests that the standard error of the values for each determinand can be used as a measure of the reconstruction error for that determinand. For the work presented in this thesis no account of the reconstruction error is made. Instead the averaged values are presented as an indication at best of the hydrochemistry likely to have existed prior to the onset of acidification, and are currently more useful in comparing the average modern analogue hydrochemistry with the present day measured chemistry in the UKAWMN lakes. The intention is not to provide hydrochemical reconstructions for the UKAWMN lakes, rather to identify biological

analogues that might provide the basis for restoration or recovery targets and a reference against which future recovery can be measured.

5.4.2.5 Lochnagar

Lochnagar is a high altitude (785 m a.s.l.) corrie loch in the Grampian Mountains of northeast Scotland. The loch is currently acidic (mean pH = 5.3) and has probably acidified from *ca.* pH 5.8 in the mid 19th Century to the present day pH, as determined by the SWAP diatom calibration set (Monteith and Evans 2000).

Loch Toll an Lochain and Llyn Clyd are the two close modern analogues identified for Lochnagar, a high altitude loch (785 m a.s.l.) in the Grampian mountains of northeast Scotland. Lochnagar has acidified from around pH 5.6 in the mid nineteenth century to pH 5.0 at the present day as indicated through diatom-inferred pH reconstruction (Patrick *et al.* 1995). Like Lochnagar, the two close modern analogues are also high altitude corrie lakes; Loch Toll an Lochain in northwest Scotland lies at 517 m a.s.l. whilst Llyn Clyd, in Snowdonia, north Wales, lies at 660 m a.s.l.

This aspect is reflected in the cladoceran fauna of the three lakes where high proportions of *A. harpae* and *A. elongata* are found, with these two taxa accounting for between 12 and 28 % of the sample total. Planktonic Cladocera are less dominant in these sites, with *B. coregoni* and *B. longispina* being found in varying proportions across the three samples, but with a total for *Bosmina* of *ca.* 50% of the sample total. *C. piger* is also found in all three of the lakes at *ca.* 6% abundance. Furthermore, the species of Cladocera found in the pre-disturbance assemblage from Lochnagar are all well represented in the training set and occur at a number of other locations. There are no species of Cladocera in the sediment sample from Lochnagar that are considered rare in the training set. This is reflected in the high number of close modern analogues (33) identified for Lochnagar when using the cladoceran training set on its own (Table 46).

Achnanthes scotica is particularly abundant in the pre-acidification sediment sample from Lochnagar, making up al most 11% of the diatom sub-fossil community of the entire sample. *A. scotica* is the dominant taxon in the sample from Lochnagar but is rare within the context of the training set. Indeed, it is found in relatively few of the 163 lakes for

which diatom data exist and only reaches a maximum abundance of 3.8%. It is not surprising therefore to find that the diatom flora found in Lochnagar prior to the onset of acidification at the site in the mid nineteenth century is quite dissimilar to the range of upland diatom floras found in the 163 lake training set as shown in Table 46. *A. scotica* is present in both of the modern analogues for Lochnagar, but only at low percentages *ca.* 1-2% of the total diatom sub-fossil community.

The other main components of the diatom flora from Lochnagar are *Fragilaria virescens* var. *exigua* and *A. minutissima.* These two taxa are also very common in the diatom floras of Loch Toll an Lochain and Llyn Clyd, as is the taxon *B. vitrea*, which is also present in all three samples though in much greater proportions in Llyn Clyd (*ca.* 14%) than the other two lakes.

Lochnagar and Loch Toll an Lochain are very similar in terms of their species diversity, richness and evenness. Llyn Clyd is not as similar in these respects, and does not contain as many species of diatom or Cladocera as either Loch Toll an Lochain or the pre-acidification assemblage of Lochnagar.

Hydrochemically, as is with the analogues for the other UKAWMN lakes discussed previously, the modern analogues are similar to the present day measured hydrochemistry of Lochnagar for the major ions and total organic carbon content of the lake water. The hydrochemistry for Loch Toll an Lochain suggests that it is much more affected by inputs of sea salts, indicated by the higher levels of chloride and sodium than those measured at Lochnagar or Llyn Clyd. This is not surprising given the location of Loch Toll an Lochain on the northwest coast of Scotland.

The modern analogues suggest a dissimilarity-weighted mean pH of 6.07 for the pre-acidification period in Lochnagar (Table 51). This is approximately 0.4 of a pH unit higher than the diatom-inferred reconstruction would suggest (Patrick *et al.* 1995) and if this value were to be believed would indicate a greater amount of change had occurred at Lochnagar as a result of acid deposition than may have been previously thought. A much higher alkalinity is suggested of 24.42 μ eq l^{-1}, which is twice as great as the maximum value recorded in Lochnagar between 1988 and 1998.

The two modern analogues identified for Lochnagar are floristically and faunistically similar, both to the pre-acidification sample and to each other. They are both high altitude corrie lakes as is Lochnagar and the inferred hydrochemistry would suggest that Lochnagar was less acidic and better buffered than the present day with a pH around 6 and strongly positive alkalinity. However, two uncertainties exist regarding the suitability of Loch Toll an Lochain and Llyn Clyd as modern analogues for Lochnagar. The first is the *ca.* 0.4 pH difference in the diatom-inferred pH reconstruction (5.6) and the dissimilarity-weighted value (6.07) inferred from the modern analogues. The second is with regard to the proportion of *A. scotica* in Lochnagar and that no close modern diatom analogues were identified using the 163-lake diatom training set (Table 46). Without further investigating the sedimentary record of Lochnagar and comparing it with the contemporary fauna and flora of Loch Toll an Lochain and Llyn Clyd it is impossible to determine the importance of these two doubts and with them whether the two lakes identified are indeed suitable modern analogues for the pre-acidification period in Lochnagar.

5.4.2.6 Llyn cwm Mynach

Llyn cwm Mynach is a small upland lake in the Rhinog Mountains in north Wales lying at an altitude of 285 m a.s.l. The lake consists of two distinct basins; the southern basin with a uniform depth of less than one metre, and the deeper northern basin which has a maximum depth of eleven metres. Palaeolimnological evidence indicates that Llyn cwm Mynach began to acidify as late as the turn of the 19th Century, from a pre-acidification pH of 5.8 to a present day pH of 5.4 (Patrick *et al.* 1995).

Table 50 lists the close modern analogues for Llyn cwm Mynach along with their mean hydrochemistry and squared chord dissimilarity with the pre-disturbance sediment sample from Llyn cwm Mynach. The mean and extreme values of the range of hydrochemical determinands for Llyn cwm Mynach recorded over the period 1988-1998 are presented.

With this number of close modern analogues, it is not surprising that there is considerably more variation in the hydrochemistry of the ten close modern analogues for Llyn cwm Mynach than when compared to the variation exhibited for Lochnagar, which has just two close modern analogues for its pre-acidification sample in the training set. The range in pH of the analogues is from 5.53 to 6.30 with a dissimilarity weighted mean pH of 5.79

and a mean pH of 5.80. This compares favourably with the present day measured pH with a maximum value of 6.30 measured between 1988 and 1998 in Llyn cwm Mynach. The mean recorded pH for Llyn cwm Mynach is 5.37 compared to the mean pH of the analogues of 5.79-5.80. This suggests that prior to acidification annual mean pH in Llyn cwm Mynach was about 0.4 pH units greater than the present day. The inferred pH value from the modern analogues compares favourably with the SWAP diatom inferred pH, which also suggests a reference state pH of 5.8. This decrease in pH over the period of acidification is accompanied by an increase in labile aluminium concentrations in the lake-water from a mean inferred concentration of 3.89 μ eq l^{-1} prior to acidification to a present day mean of 65.2 μ eq l^{-1}.

The major ionic content of the lake water from Llyn cwm Mynach is very similar to that of the majority of the close modern analogues in Table 50. The main exception to this rule is Loch Teanga (TEAN), a moderately large, low lying loch on South Uist, Outer Hebrides, Scotland. Loch Teanga has a much greater chloride and sodium concentration of 1160 μ eq l^{-1} and 981 μ eq l^{-1} respectively. These values reflect the proximity of Loch Teanga to the sea and the prevailing westerly winds. Lochan an Dubha (DUBH) also has much higher chloride and sodium concentrations than those measured between 1988 and 1998 in Llyn cwm Mynach.

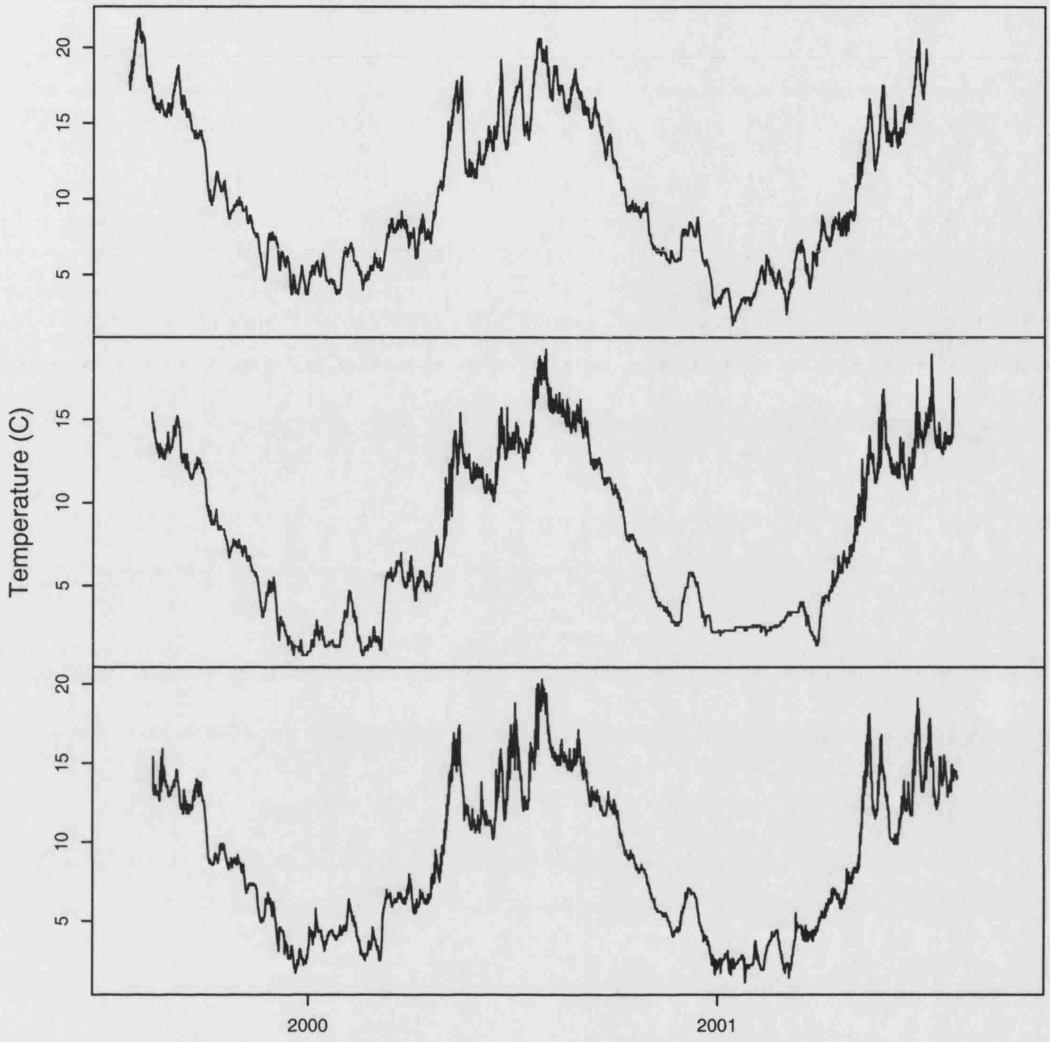


Figure 88: Plot showing time series data from Llyn cwm Mynach (top), Loch Tinker (middle) and Loch Coire nan Arr (bottom) of 2-hourly lake surface temperature recordings from late 1999 to mid 2001. The data were recorded by a data logger attached to a sediment trap located in the deepest area of each lake.

There is also a wide variation in the calcium content of the analogue lakes from 30.75 μ eq Ca l^{-1} in Loch na Eun (NEUN) to 110 μ eq Ca l^{-1} in Loch Teanga and Lochan Lairig Cheile (LACH). This variation is not as significant as it might first appear. The range of calcium values found in Llyn cwm Mynach between 1988 and 1998 is 21.5 μ eq l^{-1} to 128 μ eq l^{-1} (Monteith and Evans 2000). So, even the high values of 110 μ eq Ca l^{-1} found in Loch Teanga and Lochan Lairig Cheile are lower than the maximum record value in Llyn cwm Mynach. Furthermore, it is important to remember that at best the hydrochemical data presented in Table 50 are a mean of 4 quarterly samples, whereas the measured values in Llyn cwm Mynach represent the range of values present in monthly samples throughout the period 1988-1998. Whilst the hydrochemistry of Llyn cwm Mynach is well known as a result of the routine monitoring of the lake as part of the UKAWMN, the hydrochemistry

of the modern analogues is less well known. A previous sample from Loch Teanga (Flower *et al.* 1993) shows a lower calcium concentration of 93 μ eq l^{-1} for the loch. It is likely that, just as in Llyn cwm Mynach which has a range of over 100 μ eq l^{-1}, the two close modern analogues with high calcium concentrations also exhibit a wide range in levels of calcium in the lake water throughout the year.

The diatom and Cladocera species composition of the modern analogues is more much consistent across the range of samples than the hydrochemistry. The cladoceran fauna of the modern analogues is in the main very diverse with 5 of the modern analogues having twenty or more species of Cladocera present in the surface sediment sample. Of the remaining samples only the surface sample from Loch Tinker has fewer than fifteen species present. As one would expect, given this level of diversity in the cladoceran fauna, the planktonic species *B. coregoni* and *B. longispina* do not dominate the cladoceran assemblage unlike many samples in the training set. Instead a richer chydorid fauna is found, including a number of large bodied Cladocera (*E. lamellatus*, *S. crystallina* and *Daphnia*), in the majority of the close modern analogues. *M. dispar* and *G. testudinaria* are present in the majority of the close modern analogues as are *A. excisa*, *A. nana* and *C. piger*; the latter three species being found consistently across the range of modern analogues and often contributing 10% or more to the cladoceran composition.

The main components of the diatom flora in the modern analogues are high proportions of *B. vitrea*, *F. virescens* var. *exigua* and *F. rhomboides* var. *saxonica*. *T. flocculosa* var. *flocculosa* and *A. minutissima* var. *minutissima* are also present in many of the modern analogues. The proportion of *A. minutissima* var. *minutissima* declines as the dissimilarity to the pre-acidification sample from Llyn cwm Mynach increases, with the proportion of *E. incisa* increasing concomitantly to around 10% of the assemblage.

Like Llyn Llagi, two of the analogues for Llyn cwm Mynach are lakes in the UKAWMN; Loch Tinker from the Trossachs region of Scotland and Loch Coire nan Arr in northwest Scotland. Figure 88 shows the surface water temperatures recorded by sub-surface thermistors located on sediment traps in each of the three lakes between summer 1999 and summer 2001. The pattern of temperature changes in each of the three lakes is almost identical, but more importantly the absolute temperature values for the three lakes are very similar indeed. Despite the geographical separation between Llyn cwm Mynach in Wales

and these two Scottish modern analogues there does not appear to be any significant differences which might suggests restrictions on certain aspects of the flora and fauna between lakes.

The monitoring data available on the aquatic macrophyte flora of Loch Coire nan Arr and Loch Tinker are very similar in include three acid sensitive species often absent from acidified surface waters; *Callitriche hamulata*, *Nitella flexilis* and *Myriophyllum alterniflorum*. Unlike low growing species, such as *Littorella uniflora* and *Isoetes lacustris*, these elodeid forms grow up through the water column. These three species are also fine-leaved in contrast with the broad-leaved species like *Isoetes* and *Littorella*. These elodeid forms have particular importance for the availability of a range of habitat forms suitable for colonisation by littoral zooplankton where the structure of the littoral macrophyte community is essential in determining species composition in the littoral zone (e.g. Quade 1969; Whiteside, Williams, and White 1978).

Of these three species of acid sensitive aquatic macrophytes only *Myriophyllum alterniflorum* is currently present in Llyn cwm Mynach. *Nitella flexilis* was recorded in small numbers in 1988 at the start of the monitoring period though has not be found in the lake since. From the results of the analogue matching and the available monitoring data it would seem likely that the reference state of Llyn cwm Mynach was slightly less acid than present day conditions by almost half of one pH unit and that the lake supported a slightly more diverse aquatic macrophyte community containing a greater number of acid sensitive species than in the present day.

5.4.2.7 Round Loch of Glenhead

The Round Loch of Glenhead is a relatively small, acidic moorland loch in the Galloway Hills, southwest Scotland. The loch's catchment is largely blanket peat with *Molinia caerulea*, *Erica* spp. and *Trichophorum cespitosum* dominating the moorland vegetation. Palaeolimnological data indicate that the loch has become severely acidified with the diatom-inferred pH declining rapidly from a pre-acidification pH of *ca.* 5.3 to a pH of 4.7 by the 1960s (Monteith and Evans 2000). Present day mean measured pH at the loch is 4.9 (Monteith and Evans 2000).

The analogue matching procedure identified nine close modern analogues in the diatom and Cladocera training set. Table 52 lists these analogues alongside the mean measured chemistry for each analogue lake and summary chemistry data for the Round Loch of Glenhead over the period 1988 to 1998 from the UKAWMN.

The analogues suggest that the pre-acidification period in the Round Loch of Glenhead was much less acid than the present day. The minimum pH indicated by the modern analogues is 5.57 (Llyn Hir, HIR2), with a maximum inferred pH of 6.30 (Loch Coire nan Arr, ARR). The mean pH of the analogues and the dissimilarity-weighted mean pH both indicate a pre-acidification pH of 5.83, which is *ca.* 0.6 pH units higher than the maximum measured pH over the last ten years (= 5.21). Using the mean or weighted mean of the *n* modern analogues is an attempt to reduce the effects of natural between lake variability in chemistry where similar species occur and of variability in the chemistry not captured in the 4 quarterly samples used to make the inferences. Furthermore, the weighted mean allows one to give more importance to modern analogues most similar to the pre-disturbance sample when determining the mean inferred chemistry; this approach is justified (ter Braak 1995). The analogue inferred pH of 5.83 for the core bottom sample from the Round Loch of Glenhead is significantly higher than the diatom-inferred pH of 5.3 calculated using the SWAP calibration set (Monteith and Evans 2000). The SWAP calibration set is known to under estimate pH at the low end of the scale because of the truncation of the pH response curves for acid tolerant taxa leading to these taxa having modelled pH optima lower than the measured pH range in the training set. It could be argued, therefore, that the analogue inferred pH represents a more-likely estimate of the pH of the reference state for the Round Loch of Glenhead.

Coupled with the higher reference state pH inferred for the Round Loch of Glenhead, the analogues also suggest that the loch used to have a much higher alkalinity than the present day of between 22.72 and 23.04 μ eq l^{-1} and that aluminium levels were significantly reduced compared to the present day measured levels.

The Round Loch of Glenhead was one of the two lochs at which the original analogue matching application was applied (Flower *et al.* 1997). Loch Teanga in the Hebrides, Scotland was the closest modern analogue identified for reference state of the Round Loch of Glenhead on the basis of the diatom flora in the two samples. Loch Teanga is a much

better buffered loch than the Round Loch of Glenhead and has a mean measured calcium concentration of 100 μ eq l^{-1}. Using the joint diatom and cladoceran analogue matching procedure Loch Teanga is no longer considered a close modern analogue. Indeed, all except three of the nine close modern analogues have a mean calcium concentration of less than 50 μ eq l^{-1}. This compares well with the measured contemporary calcium concentrations (mean = 33.0 μ eq l^{-1}, max = 42.0 μ eq l^{-1}, min = 25.0 μ eq l^{-1}) though this value is still likely to be an over estimation of the calcium concentrations in the Round Loch of Glenhead prior to the onset of acidification because as the pH of the loch improves, in the short term at least, the calcium concentrations in the water will decline (Henriksen *et al.* 1992). However, the discrepancy in the calcium concentrations is now at a much more acceptable level.

An interesting observation is that Lochan Lairig Cheile (LACH) has been identified as a close modern analogue for the Round Loch of Glenhead. LACH has a present day mean measured calcium concentration of 110 μ eq l^{-1}. Llyn Hir (HIR) is also a close modern analogue for the Round Loch of Glenhead, but it too has a particularly high calcium concentration of 74 μ eq l^{-1} though this value is artificially high as the lake was limed in 1985 (Underwood, Donald, and Stoner 1987). Whilst it is difficult to generalise from four spot samples it appears that for at least two groups of species, the diatom and the Cladocera, that the same community composition can occur over a wide range of calcium concentrations. How significant this range of calcium concentrations is for other species groups still remains to be seen and work is currently ongoing to assess the similarity across the analogues for each of the UKAWMN lakes for a wider range of biological groups, including aquatic macrophytes, invertebrates and fish.

Despite the high calcium concentrations recorded for two of the analogues for the Round Loch of Glenhead the remaining analogues all have low calcium concentrations indicate an improvement in the matching routine.

The close modern analogues are very similar in terms of the main species of diatom and Cladocera to those found in the core bottom sample from the Round Loch of Glenhead. There is a close agreement between the species diversity, richness and evenness of the fossil sample from the Round Loch of Glenhead and the samples from the close modern analogues (

Table 55). The diatom flora of the pre-acidification sample from the Round Loch of Glenhead is characterised by the high proportions of acid sensitive taxa such as *A. minutissima* and *B. vitrea*, and low proportions of acidophilous and acidobiontic taxa such as *E. incisa* and *T. quadriseptata.* These taxa are consistently found in similar proportions in each of the close modern analogues. There are some exceptions, such as Loch Clair (CLAI) and Loch Lairig Cheile (LACH), which have slightly higher proportions of *E. incisa* than found in the fossil sample from the Round Loch of Glenhead, but the diatom flora of these two lochs do not contain any *T. quadriseptata.* These slight differences in the proportions of taxa do not contribute significantly to the dissimilarity between these two lochs and the reference state for the Round Loch of Glenhead.

The cladoceran fauna of the Round Loch of Glenhead and the modern analogues is typified by a large planktonic component of *Bosmina* contributing *ca.* 40% or more of the total cladoceran assemblage. Despite this large planktonic component a high diversity of chydorid taxa are also present in the samples, such as *A. harpae*, *A. nana* and *C. piger.* The only notable exception to this pattern is Llyn Hir (HIR), which has a much greater diversity of chydorids than many of the other analogues and the sample from the Round Loch of Glenhead, and a much less dominant planktonic Cladocera component.

As with Llyn Llagi and Llyn cwm Mynach, Loch Coire nan Arr is amongst the close modern analogues identified for the reference conditions of the Round Loch of Glenhead. The measured surface water temperature data at the two lochs indicate that the two lochs follow a similar thermal regime despite Loch Coire nan Arr being located further north and at an altitude almost 170 metres lower than the Round Loch of Glenhead (Figure 89).

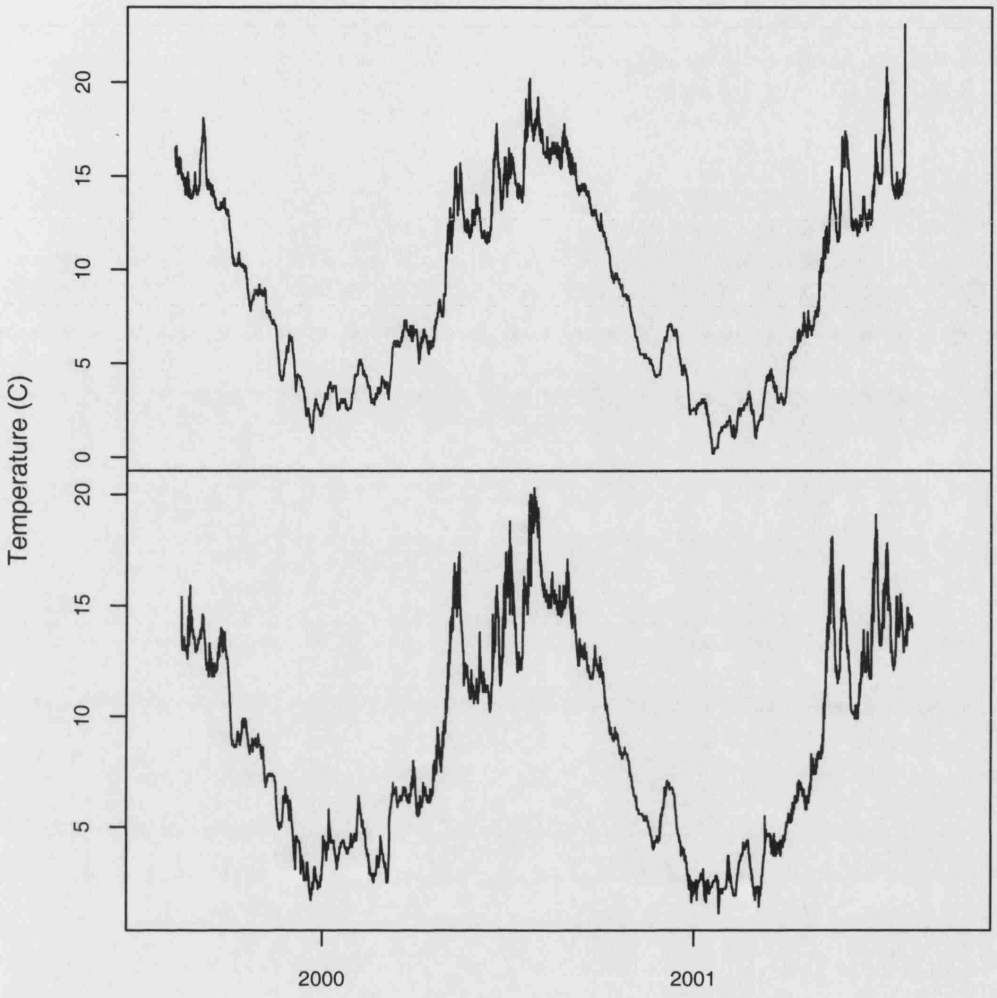


Figure 89: Plot showing time series data from the Round Loch of Glenhead (top) and Loch Coire nan Arr (bottom) of 2-hourly lake surface temperature recordings from late 1999 to mid 2001. The data were recorded by a data logger attached to a sediment trap located in the deepest area of each lake.

The aquatic macrophyte records for Loch Coire nan Arr suggest that the pre-acidification macrophyte flora in the Round Loch of Glenhead was much more divers than the present day. The Round Loch of Glenhead today has an aquatic macrophyte flora typical of acidified lakes in the UK, composed of the aquatic form of *Juncus bulbosus* and low growing isoetid species such as *Isoetes lacustris* and *Littorella uniflora* and *Lobelia dortmanna.* This flora is lacking many of the elodeid forms commonly found in non-acidified but acid sensitive upland lakes in the UK such as *Callitriche hamulata*, *Myriophyllum alterniflorum* and *Nitella flexilis.* All three of these acid sensitive species are present in Loch Coire nan Arr. It is likely that as the Round Loch of Glenhead acidified the isoetid species persisted because of their ability to derive carbon dioxide from the sediments whilst the acid sensitive species, which derive their inorganic carbon from the water column, declined as the availability of

inorganic carbon or bicarbonate in the water decreased (Smolders, Lucassen, and Roelofs 2002; Madsen, Olesen, and Bagger 2002). The high diversity of chydorid taxa present in the modern analogue lakes is indicative of a well developed and structured littoral habitat consisting of a variety of macrophyte and substrate habitats, a feature which would further indicate that a much more diverse and structured aquatic macrophyte community used to be present in the Round Loch of Glenhead prior to the onset of acidification.

5.4.2.8 Loch Tinker

Loch Tinker is a medium sized, acidic upland loch lying at an altitude of 420 m a.s.l. in the Trossachs region of central Scotland. The catchment of the loch is characterised by blanket peat overlain by *Calluna* and *Molinia* moorland vegetation. A diatom-inferred pH reconstruction for the loch suggests that it acidified from pH 6.5 in the mid 19th Century to pH 5.3 by 1913. The pH of the loch improved after 1913 to pH 5.5 by the 1970s and has remained fairly constant till the present day (Monteith and Evans 2000).

Seven close modern analogues were identified for the reference state of Loch Tinker using the joint diatom and Cladocera training set. These analogues are listed in Table 54, alongside the mean measured chemistry for each analogue lake and summary chemistry data for the Loch Tinker over the period 1988 to 1998 from the UKAWMN.

Interestingly, the surface sediment diatom and Cladocera assemblage from Loch Tinker is the closest modern analogue for the reference state of the loch. The dissimilarity between the fossil and surface samples is very low indeed indicating a high degree of similarity in both the species composition and the proportions of the main taxa in the two samples. This would suggest that the reference state for Loch Tinker is not very different from the conditions in the loch today.

The chemistry inferred from the modern analogues in general concurs with this hypothesis. The analogue-inferred pH is somewhat lower than should be expected when compared to the present day measured pH (mean = 6.13 compared to the analogue inferred pH of 5.44-5.45). This low pH is driven by the pH of three analogues (Loch na h'Achlaise [ACH], Lochan Fhionnlaidh [FHIO] and Loch na Gruagaich [GRUA]) all with a pH below 5.25. The remaining four analogues all have pH of above 5.5. Calcium concentrations are, in general,

a little low across the range of analogues where three of the analogues have a calcium concentration lower than the minimum recorded value at Loch Tinker between 1988 and 1998. Despite these two inconsistencies the chemistry for the modern analogues is in general agreement with the measured chemistry in Loch Tinker.

The species diversity, richness and evenness for the modern analogues show a close degree of agreement with that of the core bottom sample from Loch Tinker. The main diatom taxa in the fossil sample from Loch Tinker are acid sensitive taxa, such as *B. vitrea* (21%) and *Fragilaria virescens* (15.6%), and acidophilous taxa, such as *Frustulia rhomboides* var. *saxonica* (15.6%) and *E. incisa* (4.3%). These same species are also the dominant taxa in the close modern analogues and the proportions of these taxa are clearly important in determining the species composition of the sample and the similarity with the close modern analogues.

The cladoceran fauna of the core bottom sample from Loch Tinker is characterised by a high proportion of the planktonic taxon *B. longispina* and a low diversity of chydorid taxa. *Alonella excisa*, *A. harpae*, *C. piger* and *A. affinis* are the important chydorid species in the sample. The general pattern in the modern analogues is that they contain a somewhat richer chydorid community than the fossil sample from Loch Tinker. However, the main species in the fossil sample are also those chydorids that are important in the modern analogue samples, and the higher diversity of these samples is generally attributable to an occasional occurrence of a few taxa that were not recorded for Loch Tinker.

It is difficult to interpret the results of the analogue matching for Loch Tinker. Despite the SWAP diatom inferred pH reconstruction indicating that the site has acidified a surface sample from Loch Tinker is the closest modern analogue in the diatom and Cladocera training set for the pre-acidification period in Loch Tinker. Yet the 4 quarterly water chemistry samples that are associated with this surface sample from Loch Tinker would indicate a mean pH some 0.4 pH units lower than the mean contemporary pH measured at the Loch between 1988 and 1998, suggesting that the Loch was more acidic in the past than today. If the diatom inferred pH reconstruction is to be believed then Loch Tinker was recovering from a minimum pH of 5.3 since 1913, a period during which emissions of sulphur dioxide to the atmosphere were increasing.

Despite this confusion it is clear from the UKAWMN data that Loch Tinker is an example of an unpolluted, acid sensitive upland loch, with a diverse diatom, cladoceran and aquatic macrophyte community that contain many acid sensitive taxa. It is likely that the reference state for Loch Tinker is much the same as the present day and that use of other modern analogues is perhaps not, therefore, appropriate in this case.

5.5 Conclusions

The discussion in the previous section has focussed on the evaluation of the results of the analogue matching for the UKAWMN lakes. Of the ten lakes for which the new analogue matching procedure was applied eight had close modern analogues for diatoms and Cladocera in the training set. Burnmoor Tarn and Loch Chon were the two lakes for which no close modern analogues could be identified.

Three of the UKAWMN lakes (Blue Lough, Llyn Llagi and Scoat Tarn) have only one close modern analogue in the training set. Loch Coire nan Arr and Lochnagar have few close modern analogues in the training set, 3 and 2 respectively, whilst Llyn cwm Mynach, the Round Loch of Glenhead and Loch Tinker have 10, 9 and 7 close modern analogues respectively.

The lack of close modern analogues for Burnmoor and Loch Chon is probably due to the relatively small scale of the combined diatom and cladoceran training set. In the case of Burnmoor Tarn is relatively insensitive to acid deposition with a high acid neutralising capacity and Ca level. Palaeolimnological evidence indicates that the lake has not acidified despite being located in the English Lake District, a region of high sulphur deposition. This type of lake is relatively poorly represented in the training set, which has few lakes with relatively high pH and calcium (see Figure 25a). The main contributory factor to the lack of analogues for Loch Chon is the cladoceran fauna, in particular the absence of *B. coregoni* from sample. The results of the analysis of the cladoceran data presented in chapter 4 show that *B. coregoni* is a very abundant cladoceran in the training set. It is not clear why this taxon is absent from the pre-acidification sample from Loch Chon, but taxonomic problems within the *Bosmina coregoni* / *Bosmina longispina* group could be a contributory cause. *Bosmina* exhibit a large degree of morphological plasticity (Hofmann 1984) and it is possible that the taxon found in the core sample from Loch Chon is a local

morphotype. Further taxonomic and analytical work is therefore required to provide a complete analysis of the sample from Loch Chon.

For the UKAWMN lakes for which close modern analogues are found in the diatom and cladoceran training set a consistent pattern is seen in the hydrochemical inferences for the reference states of these lakes. The modern analogues suggest that conditions in most of the UKAWMN were less acidic than at the present day, with lower levels of labile aluminium and higher alkalinities also inferred. Exceptions to this rule are Loch Coire nan Arr and Loch Tinker. Surface sediment samples for these two lochs are the closest modern analogues in the training set for the core sample. This would suggest that the two lakes are at worst minimally impacted from acid deposition and, as such, the concept of a modern analogue for a largely undisturbed lake is perhaps of little relevance.

The majority of close modern analogues identified have similar lake water Ca concentrations to the ten year measured values in the UKAWMN lakes with which they have been matched. The main exception to this is for Llyn cwm Mynach, which has two close modern analogues, Lochan Lairig Cheile and Loch Teanga, which have calcium concentrations much higher than those measured in Llyn cwm Mynach between 1988 and 1998. There is a wide degree of variation in the inferred values for Ca for Llyn cwm Mynach, from 30.75 μ eq l^{-1} in Loch an Eun to 100 μ eq l^{-1} in Lochan Lairig Cheile and Loch Teanga, with the majority of the ten close modern analogues having similar lake water calcium concentrations to the measured values in Llyn cwm Mynach.

Of the two lochs studied by Flower *et al.* (1997) only the Round Loch of Glenhead was also studied in this thesis. Loch Teanga was the closest modern analogue identified using the diatom-based approach (Flower *et al.* 1997). Loch Teanga has a lake water calcium concentration that is much higher than that measured in the Round Loch of Glenhead between 1988 and 1998. The close modern analogues for the Round Loch of Glenhead identified using the diatom and cladoceran-based approach do not contain Loch Teanga. The majority of these analogues have very similar lake water calcium concentrations to the measured values for the Round Loch of Glenhead. The exceptions are Lochan Lairig Cheile (110 μ eq l^{-1}) and Llyn Hir (75 μ eq l^{-1}). It seems clear that despite Ca being an important factor in determining the distributions of the Cladocera and many other fresh water organisms in acidic lakes a similar diatom and cladoceran community can exist at

both relatively low and relatively high concentrations of *Ca*. Further work is required to assess whether the two close analogues with high Ca concentrations are as similar in terms of aquatic macrophytes, zooplankton and invertebrates as the diatom and cladoceran data would suggest.

Loch Coire nan Arr was amongst the close modern analogues identified for three of the other UKAWMN lakes (Llyn Llagi, the Round Loch of Glenhead and Llyn cwm Mynach) whilst Loch Tinker was amongst the close modern analogues for Llyn cwm Mynach. The monitoring data from the UKAWMN for these two analogues allows further inferences to be made about the reference states of these three lakes. The aquatic macrophyte data in particular suggest that the reference states of these three UKAWMN lakes had abundant stands of *Nitella flexilis*, *Myriophyllum alterniflorum* and *Callitriche hamulata*, three acid sensitive species. The presence of these three macrophytes is important because of the role they have in structuring the habitat availability in a lake. The three species are fine-leaved and grow up through the water column, unlike the *Isoetes/Lobelia/Littorella* macrophyte community common in acidified lakes in the UK. These results suggest that although the exact species composition of the modern analogues might differ from the reference state of a lake, the habitat structure of the analogue is likely to be present.

The results of the analogue matching presented in this chapter show that the modern analogue approach is a simple but powerful method of identifying reference modern analogues for acidified lakes in the UK, which requires further investigation and extending to cover a wider range of site-types encountered in acid-sensitive areas of the country. There are some outstanding issues with the technique but in general the combined diatom and Cladocera-based approach shows an improvement over the results of the diatom-based approach of Flower *et al.* (1997).

Chapter 6: Discussion and conclusions

6.1 Introduction

This chapter presents a general examination of the results of the analyses discussed in this thesis. The implications of these results for the modern analogue techniques are briefly addressed and are followed by a discussion of the methodological issues with the analogue matching approach raised in the thesis. Potential improvements to the approach are made and the possibilities for future work are discussed.

6.2 Summary of results

Chapter 3 presented an analysis of a UK-based diatom surface sediment training set. This training set, the UKAWDDS, was created by the addition of a range of recently sampled upland fresh waters from northwest Scotland (Allott and Rose 1994; Allott *et al.* 1995) and Wales (Allott and Monteith 1999) to the UK lakes included in the SWAP calibration set. The main reason for creating this new training set was to increase the extent of the environmental gradients in the data so that a wider range of lake-types was sampled. This is an important consideration because one of the issues raised by the work of Flower *et al.* (1997)was the restricted range of sites and the low representation of potential analogue lakes in the training set used for diatom-based analogue matching.

Multivariate statistical analysis showed that the main patterns in the diatom data in the UKAWDDS are explained primarily by pH and associated variables (e.g. calcium, aluminium and alkalinity). Aluminium was show to have an independent effect on the diatom distributions in the data that is uncorrelated with the covariance of aluminium with pH. The long conductivity gradient in the hydrochemical data was also particularly important in

explaining the distributions of the diatom taxa, as were maximum lake depth, altitude and TOC. The patterns evident in the diatom data of the UKAWDDS are similar to those in the diatom data from the SWAP calibration set, despite the restriction in geographical range in the UKAWDDS compared to the northwest European-coverage SWAP.

A subset of the lakes in the UKAWDDS were analysed for Cladocera remains in the surface sediments from the same samples used in the diatom analysis. 83-lakes from the UKAWDDS had sufficient sediment remaining for the cladoceran analysis. The physico-chemical properties of the subset were shown to be similar to those of the larger UKAWDDS through multivariate analyses (chapter 4).

Maximum lake depth and calcium were identified as being the most important variables in explaining variation in the Cladocera abundance data. These findings are in agreement with other studies that show that lake depth is particularly important for the Cladocera, especially the littoral-dwelling Chydoridae. Calcium is also well known to mitigate the effects of low pH and ion imbalance in acidic waters. Net relief and maximum altitude in the catchment, TOC and magnesium were also shown to be important in explaining significant amounts of variance in the distributions of the Cladocera taxa. pH did not have a significant effect on the Cladocera after the effects of other variables were accounted for.

Gaussian logit regression (GLR) models of species response curves for selected environmental variables were generated to investigate individual species' environmental requirements. A range of skewed, unimodal, sigmoidal, linear and non-significant trends was identified in the response of taxa to the environmental data. pH was shown to have an effect on a few species (*Alona guttata* var. *tuberculata*, *Alona intermedia*, *Monospilus dispar* and *Sida crystallina*) though the presence/absence of the majority of species bore little relation to pH. Maximum lake depth and lake altitude were particularly important in describing patterns in the presence/absence of a number of Cladocera taxa (e.g. *A. elongata*, *A. intermedia*, *G. testudinaria* and *S. crystallina*). The response of the these species to the two variables shows agreement between published results (e.g. Whiteside 1970; Korhola 1999) and the Cladocera data generated as part of this thesis.

Cluster analysis of the Cladocera data identified three clusters of lakes: lakes dominated by either *Bosmina coregoni* or *Bosmina longispina*, or lakes dominated by chydorid taxa. Linear

discriminant analysis of these *a priori* defined groups showed that lake and catchment areas were particularly important in explaining the differences between the two groups of lakes dominated by *Bosmina* and the group dominated by chydorid taxa. Magnesium, chloride and TOC were also important in explaining the between-group variance. The difference between the *B. coregoni* dominated lakes and those dominated by *B. longispina* was best explained by conductivity, calcium, potassium, chloride and sulphate. Maximum lake depth was also important in discriminating between these two groups of planktonic cladoceran dominated lakes.

Chapter 5 of the thesis presented the results of the analogue matching for ten of the UKAWMN lakes. The enlarged diatom training set showed little improvement in the matching over that presented by Flower *et al.* (1997). The matches produced by the combined diatom and Cladocera-based approach did, however, show considerable improvements in the appropriateness of the selected modern analogue lakes. Eight of the ten lakes had close modern analogues in the 83-lake training set using identified the combined diatom and Cladocera-based approach. The majority of these matches suggested that the reference states of the UKAWMN lakes were much less acid than at the present day and suggested that future recovery targets would be lower aluminium levels and improvements in alkalinity. Calcium concentrations between the close modern analogues and the measured data for the UKAWMN lakes showed a high degree of agreement, though a couple of the matches did have much higher concentrations than those measured in the lakes for which they were analogues. The main findings of this chapter were that the improved diatom and cladoceran-based analogue matching approach was better at identifying matches that the diatom-based approach, and that the new method is a simple yet reliable technique for identifying modern analogues for acidified lakes in the UK.

6.3 Implications of the results

This section of the thesis will briefly address the implications of the findings of chapters 4 and 5 for future work using Cladocera or analogue matching.

6.3.1 Implications for future Cladocera-based palaeolimnological work

The results of the analysis of the cladoceran data in chapter 4 clearly show the complexity and range of responses of the Cladocera to environmental gradients. In the acidic lakes sampled in this study, important gradients included maximum lake depth, calcium, TOC, altitude and magnesium. The twenty-four physico-chemical variables used in this analysis explain approximately half the variance in the species data. The remaining variance is likely to be related to habitat structure and availability (including type and abundance of aquatic macrophytes), and predator-prey relationships, amongst others. How all these factors interrelate is unclear and as yet all these factors have not been considered together in research work.

Despite our relatively rudimentary understanding of the nature of the responses of Cladocera to a combination of environmental factors, researchers have generated cladoceran-based transfer functions for a range of environmental variables, most notably for pH (Krause-Dellin and Steinberg 1984) and temperature (Lotter *et al.* 1997; Korhola 1999), but also for lake depth (Korhola *et al.* 2000).

Krause-Dellin and Steinberg's (1984) original relationship between Cladocera and pH is derived from a multiple regression of pH preference groups for cladoceran taxa against pH in eight lakes over a wide pH range from 3.6 to 6.7. It is not surprising, therefore, that the authors found differences in Cladocera species composition along this pH gradient. In a more-comprehensive study of Cladocera-pH relationships, the results presented in chapter 4 suggests very little in the way of a response by Cladocera to pH. Of the forty-three taxa included in the current work only four showed a significant unimodal or linear response to pH. Partial redundancy analysis of the Cladocera species data has shown that pH explains very little of the variance in the species data after the effects of other, more important variables with greater explanatory power, have been accounted for.

It is clear from palaeolimnological (e.g. Paterson 1994) and experimental manipulation work (e.g. Hann and Turner 2000) that a community wide response to acidification is seen in the Cladocera. This community response is likely to be related more to changes in aquatic macrophytes (in particular the loss or reduction of elodeid forms like *Myriophyllum alterniflorum* or *Nitella flexilis* and associated changes in the proportions of available

habitats), in food sources resulting from phytoplankton community responses to acidification (e.g. Vinebrooke *et al.* 2002), in relative predation pressures on part of the cladoceran community due to changes in invertebrate and/or fish predators, all of which are well-known whole-lake responses to acidification, rather than to the direct effects of the pH change. And because the response of the Cladocera is known to be related to such a wide range of factors that are all affected by acidification, the palaeolimnological record of Cladocera in acidified lakes should contain this information. By understanding more about the relationships between the Cladocera and these other factors, and by applying that understanding to the interpretation of the cladoceran palaeolimnological record in acidified lakes it should be possible to gain greater insights on the impacts of acidification in these systems. It is important to use cladoceran data in the correct manner. Diatom-based pH transfer functions are well developed and backed by conclusive evidence of effects of pH on diatom physiology. It is, therefore, not necessary to use the Cladocera as a means of reconstructing pH from remains in lake sediments. Research efforts would be far better spent in untangling the complex environmental responses of the Cladocera to better understand how acidification affected whole-lake ecosystems through time.

The same caution should be applied palaeolimnological studies, such as climate change, and greater emphasis should be placed on interpreting the species compositional changes in sediment cores and extracting as much information from cladoceran remains, such as changes in the number of ephippia through time (Sarmaja-Korjonen 2002) with respect to whole-lake changes rather than the whole sale application of transfer functions.

6.3.2 Implications for defining reference states for acidified lakes

The development of the analogue matching approach presented in chapter 5 clearly indicates the effectiveness of such an approach in defining recovery targets for acidified lakes. The results also demonstrate the utility of not restricting palaeolimnological research to a single proxy. The approach does not attempt to define the reference state of an acidified lake directly; rather it identifies suitable modern analogues. These modern analogues can then be studied to determine the range of taxa likely to have been present in an acidified lake prior to the onset of acidification. This two-step approach is beneficial because information on a wide range of organisms that are not preserved in lakes sediments can be collected from minimally impacted reference systems, the modern

analogue lakes. This is particularly important because little is known about the biological communities found in acidified lakes prior to the onset of acidification in the UK, save for the palaeolimnological record, as the majority of these lakes are in remote areas of the UK for which few, if any, biological records exist.

The European Union's Water Framework Directive (WFD) places particular importance on evaluating changes in the ecological status of waters and assessing how different present day conditions are from those expected in the absence of significant, recent anthropogenic impacts, such as acidification and eutrophication. Palaeolimnological techniques, and in particular analogue matching, can provide this type of information and will be essential in underpinning water management policies in the UK to meet the requirements of the WFD. The analogue matching approach has, however, only been developed for use in acidified surface waters and further work will be required to transfer the methodology developed in this thesis to other lake types such as lowland lakes, mesotrophic lakes and large, deep oligotrophic standing waters which have received little in the way of integrated, multi-proxy palaeolimnological work at present (Bennion *et al.* 2002).

6.4 Methodological issues

The research presented in chapter 5 has highlighted a number of methodological issues that, while not undermining the usefulness of the analogue matching approach for defining ecologically meaningful reference conditions, do need addressing.

One of the main issues is how to integrate distinct data sets of palaeolimnological data for different species into a single data set used for analogue matching. The method used in chapter 5 was to take the percentage count data for the diatoms and the Cladocera and to append one data set to the other and then divide all the percentage values by two so that the sample totals added up to 100. This approach is convenient as it does not explicitly give additional weighting to one or other of the species groups involved. What this method does not take into account, however, are the large differences in the number of diatom taxa present in sediment samples compared to the number of cladoceran taxa. The differences in the numbers of taxa in each of the species groups used in the analogue matching could lead to greater weight being given implicitly to the species group that has greater number of taxa. The reason why greater weight might be given to the group with

more species is that, as a proportion of the number of comparisons between the species present in any two samples, the group with the larger number of taxa will contribute more greatly than the group with fewer species.

188 taxa were present in the diatom data used in the combined diatom and cladoceran data set compared to the 47 Cladocera taxa identified. As such, the diatom taxa could be contributing four times more information to the dissimilarity calculations across the whole training set than the Cladocera. It is likely that due to this discrepancy in the numbers of taxa that the diatom data had greater importance in determining the degree of similarity between any two samples than the Cladocera data.

The differences in the number of taxa in the entire data set between the Cladocera and the diatoms do not give, however, a reliable indication of the additional influence of the diatoms in the dissimilarity calculations. The relevant influence will be sample-pair specific, with the numbers of taxa per group in each sample determining the extra importance placed upon the diatoms in the dissimilarity calculations. In the combined training set, this influence could range from about two to over four times more influence on the part of the diatoms than the Cladocera, based on an assessment of the numbers of diatom and Cladocera taxa in each sample in the training set. The actual influence then is determined by the relative number of diatom to cladoceran taxa in the fossil sample and each modern sample. As such, it is difficult to envisage the impact this additional influence might have on the matching results other than that the diatom species composition could have greater weighting in the calculation of dissimilarity on a sample by sample basis.

A further complication is that the squared chord distance measure is known to be robust to small-scale differences between samples because it is a signal-to-noise dissimilarity coefficient (Overpeck *et al.* 1985). The majority of the difference in the numbers of diatom and Cladocera taxa will be in rare diatom taxa present in small proportions. As such these taxa do not contribute as much as the more abundant taxa to the dissimilarity between two samples.

One solution to this problem might be to use the same number of diatom and Cladocera taxa in the training set used for analogue matching. The dissimilarity would then be based on the most abundant diatom species and the Cladocera taxa present. This solution would,

however, delete a great deal of information from the matching process and further work is required to assess the amount rare taxa contribute to the dissimilarity between samples and the effect of different numbers of taxa per species group on the relative importance of the groups in the dissimilarity calculations.

The second major methodological issue identified in chapter 5 is the determination of critical values for defining *close* modern analogues. Four approaches are mentioned in chapter 5; use expert knowledge to set critical values, use the modern training set to empirically set critical values based on the distribution of dissimilarity values generated by comparison of every modern sample with every other sample, permute the modern training set many times and generate a random distribution of dissimilarity values, or generate random dissimilarity data of known properties and use this distribution to generate critical values.

Of the four methodologies, the latter was used in chapter 5 as an alternative to the other three methods. Using expert knowledge was discounted as impractical because of the difficulties involved in translating the expert knowledge into numerical values that could be used as critical values for determining the level of dissimilarity required for samples to be defined as *close* modern analogues. The second method is highly influenced by the distribution of dissimilarity values in the training set, with deviations from a normal distribution leading to overly high or low critical values depending on the direction of skew. Permutation of the species data matrix to generated *n* random dissimilarity values appears to be an ideal way of determining the critical values one should use. The permutation test used in ANALOG, shuffles the data in the columns of the data matrix at random. New row totals are calculated and the data converted into new percentages based on these new row totals. In this way, values for naturally rare taxa are kept within the range of the training set and unnaturally large values for these taxa are not generated. In the case of the diatom and cladoceran data generated for thesis, the permutation test generated critical values greater than 1.0 in many cases for the 5^{th} percentile. Clearly, values this large do not indicate samples that are very similar at all, and there appears to be many more ways of permuting the data to generate dissimilar sets of samples than similar ones. By using randomly generated data whose values lie between 0 and 2 and which follow a normal distribution an alternative approach to the empirical and permutation based approaches can be found. Using the 5^{th} percentile of this distribution of randomly generated data provides an

independent assessment of the critical value of dissimilarity required for *close* analogues to be identified.

The latter of the four methodologies can only be used with those dissimilarity coefficients that take a maximum and minimum value. In the case of the squared chord distance, for example, the maximum and minimum values are 2 and 0 respectively. For data sets where the chord distance or similar measure is appropriate (i.e. where the emphasis is on the differences in the proportions of taxa between samples) then this approach may be a simple way of defining critical values and one which is transferable between studies because once the critical values have been determined from the randomly generated data they do not change as the data set changes. Where the emphasis is on different properties of the data, such as where differences in the magnitudes of the abundance of taxa between two samples are of interest a Euclidean distance might be used then this approach would not be suitable as the Euclidean distance does not assume an upper limit.

Further work is required with data of known properties and samples of known similarities to assess the level of similarity required for close analogues to be defined, and is discussed below.

6.5 Opportunities for further research

The work presented in chapter 5 describes the methodology used to identify modern analogues that can be used as reference states for presently acidified lakes and discusses the results of analogue matching using the new approach for the UKAWMN lakes.

The major piece of work that remains to be completed and assessed is a study into how well the modern analogues identified for each of the UKAWMN lakes compare to the reference states of acidified lakes in terms of their wider biological groups. This work entails using a wider range of proxies from sediment cores to compare, for example, the remains of invertebrate groups such as chironomid or chaoborid larvae with those found in the modern analogue lake in the present day. An analysis of the sediment record of acidified lakes for plant macrofossil remains could also be compared with the aquatic macrophytes found in the modern analogue lakes to add further biological groups to the

process. An analysis could then be made of the degree of similarity between the invertebrate or aquatic macrophyte communities found in the sediment records of acidified lakes and those found in the analogue lakes, and this similarity compared to the similarity indicated by the diatoms and Cladocera.

There are problems with preservation and taphonomy with this approach, however, because not all biological groups leave identifiable remains in lake sediments and because not all members of any one biological group are preserved equally as well. As a result of these problems, the scope for comparing fossil indicators in acidified lakes with the contemporary biology of the analogue lakes is limited.

An alternative approach for evaluating the results of the diatom and cladoceran-based analogue matching might be to survey the analogue lakes for a range of contemporary biological groups; aquatic macrophytes, epilithic diatoms, zooplankton and invertebrates. For those UKAWMN lakes that have more than one close modern analogue, the dissimilarity between each of the modern analogues can be assessed based on the range of groups surveyed in the analogue lakes. If the diatom and cladoceran-based analogue matching approach is producing reliable matches then the similarity in the contemporary biology of the modern analogues should be high, conversely, if the combined approach is not providing reliable matches, then we would expect this similarity to be low.

Neither of these approaches is entirely satisfactory in isolation. Comparing a wider range of proxies from sediment cores with those same proxies in the modern analogue lakes is restricted to those organisms that preserve that preserve in lake sediments, and, whilst comparisons amongst the analogues for a single acidified lake do consider a range of organisms that do not preserve in lake sediments, these comparisons do not address directly how similar the analogues are to the pre-acidification state of the acidified lake.

A combined approach is, therefore, likely to provide the best assessment of the combined analogue matching procedure. Research to address this issue is currently being pursued for the Department of Environment, Food and Rural Affairs (DEFRA) by the Environmental Change Research Centre (ECRC) and ENSIS Ltd. (DEFRA contract EPG 1/3/183), which is taking the combined approach to assessing the results of the diatom and Cladocera-based approach in twenty-eight upland lakes in Wales and Scotland.

Aside from evaluating the output of the new analogue matching approach, further work is required on the methodological issues discussed in the previous section. Furthermore, improvements to the approach itself might be justified in the light of any deficiencies highlighted by the ongoing evaluation work. Three improvements could be considered; increasing the size of the training set used in the matching process, adding further proxies to the diatoms and the Cladocera, and developing and further investigating the methodology used in the approach.

Of the ten UKAWMN lakes for which analogue matching using the diatom and Cladocera approach has been applied two had no close modern analogues in the modern training set and a further three lakes had just a single close modern analogue. The inability of the current methodology to find close modern analogues for all the lakes in the UKAWMN is due to there being insufficient suitable modern analogues in the modern training set. Additional lakes could be sampled to increase the number and range of lakes in the modern training set. Particular importance could be placed on sampling more lakes with coniferous afforestation in their catchments to address potential issues raised by the lack of analogues for Loch Chon, and a greater number of lakes above the notional tree line (500 m in Scotland) could be sampled to try to find more close modern analogues for lakes like Scoat Tarn and Lochnagar. The number of naturally very acidic lakes could also be improved to try to find analogues for lakes like Blue Lough. By improving the coverage of the modern training set the edge effects, where by any two similar samples are likely to be positioned far apart in multivariate space in at least one dimension (ter Braak 1995), are reduced and the available pool of potential analogues is increased.

The utility of adding the Cladocera to the matching process has been illustrated in chapter 5. Additional improvements might be achieved by adding further proxies to the matching process. Suitable proxies are limited, however, by the selective preservation of organisms living in surface waters. A further consideration should be to try to represent higher levels of the trophic structure of lakes in the matching process. The diatom and cladoceran-based approach presented in chapter 5 uses a member of the primary producers in lakes (diatoms) and a member of the zooplankton from a trophic level above the primary producers (Cladocera). The next obvious choice of suitable proxy is the head capsule remains of larval chironomids or non-biting midges (Walker 2001). Larval chironomid head capsules are found abundantly in lake sediments and have been used widely in

palaeoecological work, particularly in temperature reconstructions and investigations of climate change through the Holocene (e.g. Walker, Mott, and Smol 1991; Wilson *et al.* 1993; Lotter *et al.* 1997; Olander *et al.* 1999; Korhola *et al.* 2000; Vasko, Toivonen, and Korhola 2000; Korhola *et al.* 2002). By adding chironomid remains to the matching process another level of the trophic structure will be part of the matching process and could potentially produce increasingly robust matches.

Problems with the methodology used to set critical values for the definition of close modern analogues were discussed in the previous section. The importance of rare taxa in the matching process could be investigated by assessing how dissimilarity between two samples changes as taxa are removed. If this is extended to artificial data sets with different levels of noise (i.e. rare taxa) then an assessment of how similarity changes as rare taxa are excluded can be made without the complications of knowing *how similar* two real-world samples actually are. Artificial data sets could also be used to assess the effect on the measure of similarity of differences in the numbers of taxa per species group. Artificial data sets with different numbers of taxa could be created to represent different species groups, which could then be used in dissimilarity calculations to assess the importance of having similar numbers of taxa included in each of the species groups. These two pieces of work should be addressed at the same time, because one implication of using similar numbers of taxa in each of the species groups used in the matching process is that the rare taxa are excluded from the calculations for groups at lower trophic levels (e.g. the diatoms). This is of particular importance if additional or different proxies are used in the matching process because the higher up the trophic structure a group is the fewer taxa there are. For a hypothetical situation where diatoms Cladocera and chironomids are used in an analogue matching approach, many diatom taxa and some cladoceran taxa could be discounted in the dissimilarity calculations because fewer chironomid taxa were found in the samples.

The area of permutation tests to assess how similar two samples really are needs further work. The restrictions in the permutation tests in ANALOG are too weak (Birks, H.J.B., *pers. comm.*) and produce too many samples that are very dissimilar to one another; a pattern which is not seen in real, unpermuted data (see Figure 79-Figure 82). It is not immediately clear, however, as to how one might improve the restrictions in the permutation test.

A further area where there is great potential for improving and applying the analogue matching approach is the extension of the approach to a range of different lake-types, such as lowland and mesotrophic lakes that have experienced problems with anthropogenic eutrophication. The WFD calls on all members of the European Union to restore all disturbed surface waters and groundwaters to pre-disturbance targets. The same issues with defining reference states that apply to acidified surface waters, such as the lack of historical data, are equally applicable in other types of surface water. Palaeolimnological techniques, especially analogue matching, will be particularly important tools for the definition of reference conditions for a wider range of lake types. Additional work will be required, however, to assess which are the appropriate biological proxies to be used in different lake types and sampling exercises will need to be undertaken to generate large species data sets that cover enough environmental space so as to cover a range of potential analogues and to reduce the edge effects in sparse multivariate data.

There is one possible difficulty with applying analogue matching across a wide range of lake types, however, in that suitable analogues might not exist in the UK for particular lake types. Suitable analogues might be found in Europe, though biogeographical differences across the modern training sets would have to be investigated to ensure species responses to environmental gradients are similar throughout the geographical range of the training set.

6.6 Conclusions

This study comprises three main components; the analysis of a UK-based diatom training set, the creation and analysis of a cladoceran training set, and the development and application of a diatom and Cladocera-based analogue matching procedure.

The UKAWDDS illustrated similar diatom environment relationships to other training set from acid lakes (e.g. SWAP) and demonstrated the important role of pH in explaining the distributions of diatom taxa. The creation of a larger diatom-based training set for analogue matching, however, did not improve the quality of the matches identified and the problems with the original approach remained (Flower *et al.* 1997).

The results of the analysis of the Cladocera training set demonstrate the complex nature of the species-environment relationships in this group of crustacean zooplankton. Maximum depth and lake water calcium concentrations were shown to be particularly important in explaining variation in the cladoceran species data using a range of multivariate and univariate numerical techniques. Other important variables include net relief and maximum altitude in the catchment, TOC, magnesium and lake and catchment area. It was also demonstrated that the diatoms and the Cladocera respond to different environmental factors and that when combined into a single data set the properties of both species groups were retained in the combined data set.

Analogue matching using this combined data set identified suitable modern analogue lakes for eight of the ten UKAWMN lakes studied in chapter 5. These matches were shown to be suitable in terms of their hydrochemistry, indicating that the reference conditions in the acidified UKAWMN lakes were less acidic with lower aluminium levels and higher alkalinities, whilst the concentrations of the major cations were similar to the present day measured concentrations in the lakes. Additional data sources available for some of the modern analogue lakes identified using the approach confirms the reliability of the matching process.

Further research is required to validate the modern analogues identified using the combined diatom and Cladocera-based approach and to investigate the potential of adding additional biological proxies to the matching process. Methodological issues remain to be solved, in particular the importance of differences in the number of taxa per species group in the dissimilarity calculations needs to be quantified, and the development of new permutation tests to assess the significance of the similarity between potential analogues and the fossil assemblages.

The main findings of this work indicate that diatom and Cladocera-based analogue matching is a simple but powerful technique for identifying reference conditions in acidified lakes. The technique has particular relevance to the fulfilment of the requirements of the European Union's Water Framework Directive and analogue matching could potentially be applied to a wider range of lake types to this end.

List of References

Albers,P.H., Prouty,R.M. (1987) Survival of spotted salamander eggs in temporary woodland ponds of coastal Maryland. *Environmental Pollution*, **46**, 45-61.

Alibone,M.R., Fair,P. (1981) The Effects of Low pH on the Respiration of *Daphnia magna* Straus. *Hydrobiologia*, **85**, 185-188.

Allott,T.E.H., Golding,P.N.E., Harriman,R. (1995) A palaeolimnological assessment of the impacts of acid deposition on surface waters in north-west Scotland, a region of high sea-salt inputs. *Water, Air, and Soil Pollution*, **85**, 2425-2430.

Allott,T.E.H., Harriman,R., Battarbee,R.W. (1992) Reversibility of lake acidification at the Round Loch of Glenhead, Galloway, Scotland. *Environmental Pollution*, **77**, 219-225.

Allott, T. E. H. and Monteith, D. T. Classification of Lakes in Wales for Conservation using Integrated Biological Data. ECRC Research Report No. 53. 1999. London, ECRC.

Allott, T. E. H. and Rose, N. L. A palaeolimnological study of recent water quality changes in lochs with black throated diver populations. ECRC Research Report No.5. 1994. London, ECRC.

Almer,B., Dickson,W., Ekström,C., Hörnström,E., Miller,U. (1974) Effects of acidification on Swedish lakes. *Ambio*, **3**, 30-36.

Alonso,M. (1996) *Crustacea, Branchiopoda*, 1-486, Museo Nacional de Cienias Naturales. CSIC., Madrid.

Anderson,N.J. (1986) Diatom biostratigraphy and comparative core correlation within a small lake basin. *Hydrobiologia*, **143**, 105-112.

Anderson,N.J. (1989) A whole-basin diatom accumulation rate for a small eutrophic lake in Northern Ireland and its palaeoecological implications. *Journal of Ecology*, **77**, 926-946.

Anderson,N.J. (1990a) Spatial pattern of recent sediment and diatom accumulation in a small, monomictic, eutrophic lake. *Journal of Paleolimnology*, **3**, 143-160.

Anderson,N.J. (1990b) Variability of diatom concentrations and accumulation rates in sediments of a small lake basin. *Limnology and Oceanography*, **35**, 497-508.

Anderson,N.J. (1990c) Variability of sediment diatom assemblages in an upland, wind-stressed lake (Loch Fleet, Galloway, S.W. Scotland. *Journal of Paleolimnology*, **4**, 43-59.

Anderson,N.J., Battarbee,R.W. (1994) Aquatic community persistence and variability: a palaeolimnological perspective. In: *Aquatic Ecology: Scale, pattern and process. The 34th Symposium of the British Ecological Society*, 233-259, Blackwell Scientific Publications, Oxford.

Anderson, N. J., Battarbee, R. W., Appleby, P. G., Stevenson, A. C., Oldfield, F., Darley, J., and Glover, G. Palaeolimnological evidence for the acidification of Loch Fleet. Palaeoecological Research Unit Working Paper No. 17. 1986. University College London.

Anderson,N.J. (1995) Using the past to predict the future: lake sediments and the modelling of limnological disturbance. *Ecological Modelling*, **78**, 149-172.

Appleby,P.G., Nolan,P.J., Gifford,D.W., Godfrey,M.J., Oldfield,F., Anderson,N.J., Battarbee,R.W. (1986) ^{210}Pb dating by low background gamma counting. *Hydrobiologia*, **143**, 21-27.

Arzet,K., Krause-Dellin,D., Steinberg,C. (1986) Acidification of four lakes in the Federal Republic of Germany as reflected by diatom assemblages, cladoceran remains and sediment chemistry. In: *Diatoms and Lake Acidity*, 227-250, Dr W. Junk Publishers, Dordrecht.

Atkinson,K.M., Haworth,E.Y. (1990) Devoke Water and Loch Sionascaig: recent environmental changes and the post-glacial overview. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 349-355.

Barber,H.G., Haworth,E.Y. (1981) *A guide to the morphology of the diatom frustule*, 1-112, Freshwater Biological Association.

Battarbee,R.W. (1984) Diatom analysis and the acidification of lakes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **305**, 451-477.

Battarbee,R.W. (1990) The causes of lake acidification, with special reference to the role of acid deposition. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 339-347.

Battarbee,R.W. (1991) Recent paleolimnology and diatom-based environmental reconstruction. In: *Quaternary Landscapes*, 129-174, University of Minnesota Press, Minneapolis.

Battarbee,R.W. (1994) Diatoms, lake acidification and the Surface Waters Acidification Programme (SWAP): a review. *Hydrobiologia*, **274**, 1-7.

Battarbee,R.W. (1997) Freshwater Quality, Naturalness and Palaeolimnology. In: *Freshwater Quality: defining the indefinable?*, 155-171, The Stationary Office, Edinburgh.

Battarbee,R.W. (1999) The importance of palaeolimnology to lake restoration. *Hydrobiologia*, **395/396**, 149.

Battarbee,R.W., Anderson,N.J., Appleby,P.G., Flower,R.J., Fritz,S.C., Haworth,E.Y., Higgitt,S., Jones,V.J., Munro,M.A.R., Natkanski,J., Oldfield,F., Patrick,S.T., Richardson,N.G. (1988a) *Lake acidification in the UK*, 1-68, ENSIS Publishing Limited, London.

Battarbee,R.W., Charles,D.F. (1986) Diatom-based pH reconstruction studies of acid lakes in Europe and North America: a synthesis. *Water, Air, and Soil Pollution*, **30**, 347-354.

Battarbee,R.W., Flower,R.J., Stevenson,A.C., Jones,V.J., Harriman,R., Appleby,P.G. (1988b) Diatom and chemical evidence for reversibility of acidification of Scottish lochs. *Nature*, **332**, 530-532.

Battarbee,R.W., Flower,R.J., Stevenson,A.C., Rippey,B. (1985) Lake acidification in Galloway: a palaeoecological test of competing hypotheses. *Nature*, **314**, 350-352.

Battarbee,R.W., Jones,V.J., Flower,R.J., Cameron,N.G., Bennion,H., Carvalho,L., Juggins,S. (2001) Diatoms. In: *Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal , and Siliceous Indicators.*, 155-202, Kluwer Academic Publishers, Dordrecht.

Battarbee,R.W., Renberg,I. (1990) The Surface Waters Acidification Project (SWAP) Palaeolimnology Programme. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 227-232.

Beamish,R.J., Lockhart,W.L., Van Loon,J.C., Harvey,H.H. (1975) Long-term acidification of a lake and resulting effects on fishes. *Ambio*, **4**, 98-102.

Beebee,T.J.C., Flower,R.J., Stevenson,A.C., Patrick,S.T., Appleby,P.G., Fletcher,C., Marsh,C., Natkanski,J., Rippey,B., Battarbee,R.W. (1990) Decline of the natterjack toad *Bufo calamita* in Britain - Palaeoecological, documentary and experimental evidence for breeding site acidification. *Biological Conservation*, **53**, 1-20.

Bennion,H. (1994) A diatom-phosphorus transfer function for shallow, eutrophic ponds in south-east England. *Hydrobiologia*, **275/6**, 391-410.

Bennion,H., Simpson,G.L., Battarbee,R.W., Cameron,N.G., Curtis,C., Flower,R.J., Hughes,M.J., Jones,V.J., Kernan,M.R., Monteith,D.T., Patrick,S.T., Rose,N.L., Sayer,C.D., Yang,H. (2002) Environmental change in Scottish fresh waters. In: *The state of Scotland's environment and natural heritage* Scottish Natural Heritage, Edinburgh.

Berglund,B.E. (1986) *Handbook of Holocene palaeoecology and palaeohydrology*, John Wiley & Sons Ltd., UK.

Bick,H., Drews,E.F. (1973) Self purification and silicate communities in an acid milieu. *Hydrobiologia*, **42**, 393-402.

Birks,H.J.B. (1996) Environmental Change in Britain - a Long-term Palaeoecological Perspective. In: *Britain's Natural Environment: a State of the Nation Review*, 23-28, ENSIS Publications Ltd., London.

Birks,H.J.B., Juggins,S., Line,J.M. (1990a) Lake surface-water chemistry reconstructions from palaeolimnological data. In: *The Surface Waters Acidification Programme*, 301-313, Cambridge University Press, Cambridge.
Birks,H.J.B., Line,J.M., Juggins,S., Stevenson,A.C., ter Braak,C.J.F. (1990b) Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 263-278.
Blancher,P.J., McNicol,D.K. (1988) Breeding biology of tree swallows in relation to wetland acidity. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*, **66**, 842-849.
Bootsma,M.C., Barendregt,A., van Alphen,J.C.A. (1999) Effectiveness of reducing external nutrient load entering a eutrophicated shallow lake ecosystem in the Naardermeer nature reserve, The Netherlands. *Biological Conservation*, **90**, 193-201.
Borcard,D., Legendre,P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045-1055.
Bradshaw,A.D. (1984) Land restoration - now and in the future. *Proceedings of the Royal Society of London Series B Biological Sciences*, **223**, 1-23.
Bradshaw,A.D. (1996) Underlying principles of restoration. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 3-9.
Breiman,L. (1996) Bagging predictors. *Machine Learning*, **24**, 123-140.
Breiman,L. (2001) Random forests. *Machine Learning*, **45**, 5-32.
Brodin,Y.M., Gransberg,M. (1993) Responses of insects, especially Chironomidae (Diptera), and mites to 130 years of acidification in a Scottish lake. *Hydrobiologia*, **250**, 201-212.
Brouwer,E., Bobbink,R., Roelofs,J. (2002) Restoration of aquatic macrophyte vegetation in acidified and eutrophied softwater lakes: an overview. *Aquatic Botany*, **73**, 405-431.
Bull,K.R. (1995) Critical loads - Possibilities and constraints. *Water, Air, and Soil Pollution*, **85**, 201-212.
Cameron,N.G. (1995) The representation of diatom communities by fossil assemblages in a small acid lake. *Journal of Paleolimnology*, **14**, 185-223.
Cameron,N.G., Birks,H.J.B., Jones,V.J., Berge,F., Catalan,J., Flower,R.J., Garcia,J., Kawecka,B., Koinig,K.A., Marchetto,A., Sanchez-Castillo,P., Schmidt,R., Sisko,M., Solovieva,N., Stefkova,E., Toro,M. (1999) Surface-sediment and epilithic diatom pH calibration sets for remote European mountain lakes (AL : PE Project) and their comparison with the Surface Waters Acidification Programme (SWAP) calibration set. *Journal of Paleolimnology*, **22**, 291-317.
Cattell,R.B. (1966) The data box: its ordering of total resources in terms of possible relational systems. In: *Handbook of multivariate experimental psychology*, 67-128, Rand McNally & Co., Chicago.
Charles,D.F., Dixit,S.S., Cumming,B.F., Smol,J.P. (1991) Variability in diatom and chrysophyte assemblages and inferred pH: palaeolimnological studies of Big Moose Lake, New York, USA. *Journal of Paleolimnology*, **5**, 267-284.
Charles,D.F., Smol,J.P., Engstrom,D.R. (1994) Paleolimnological Approaches to Biological Monitoring. In: *Biological Monitoring of Aquatic Systems*, 233-293, CRC Press, Boca Raton, Florida.
Charles,D.F., Whitehead,D.R. (1986) The PIRLA project: palaeoecological investigations of recent lake acidification. *Hydrobiologia*, **143**, 13-20.
Cleveland,W.S., Devlin,S.J. (1988) Locally weighted regression - an approach to regression- analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596-610.
Clymo,R.S. (1984) Sphagnum-Dominated Peat Bog - A Naturally Acid Ecosystem. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **305**, 487-499.
Cosby,B.J., Ferrier,R.C., Jenkins,A., Wright,R.F. (2001) Modelling the effects of acid deposition: refinements, adjustments and inclusion of nitrogen dynamics in the MAGIC model. *Hydrology and Earth System Sciences*, **5**, 499-517.
Cosby,B.J., Hornberger,G.M., Galloway,J.N., Wright,R.F. (1985a) Modelling the effects of acid deposition: assessment of a lumped parameter model of soil water and streamwater chemistry. *Water Research*, **21**, 51-63.
Cosby,B.J., Hornberger,G.M., Galloway,J.N., Wright,R.F. (1985b) Time scales of acidification: a quantitative model for estimating freshwater acidification. *Environmental Science and Technology*, **19**, 1144-1149.
Cumming,B.F., Davey,K.A., Smol,J.P., Birks,H.J.B. (1994) When did acid-sensitive Adirondack lakes (New-York, USA) begin to acidify and are they still acidifying. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1550-1568.
Davis,R.B., Norton,S.A., Hess,C.T., Brakke,D.F. (1983) Paleolimnological reconstruction of the effects of atmospheric deposition of acids and heavy-metals on the chemistry and biology of lakes in New-England and Norway. *Hydrobiologia*, **103**, 113-123.
Deevey,E.S., Deevey,G.B.J. (1971) The American species of *Eubosmina* Seligo (Crustacea, Cladocera). *Limnology and Oceanography*, **16**, 201-218.
Delcourt,H.R., Delcourt,P.A., Webb,T.I. (1983) Dynamic plant ecology: The spectrum of vegetational change in time and space. *Quaternary Science Reviews*, **1**, 153-175.
DeMelo,R., Hebert,P.D.N. (1994a) A taxonomic reevaluation of North-American Bosminidae. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*, **72**, 1808-1825.
DeMelo,R., Hebert,P.D.N. (1994b) Allozymic variation and species-diversity in North-American Bosminidae. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 873-880.
DeMelo,R., Hebert,P.D.N. (1994c) Founder effects and geographical variation in the invading cladoceran *Bosmina* (*Eubosmina*) *coregoni* Baird 1857 in North- America. *Heredity*, **73**, 490-499.
DesGranges,J.-L., Hunter,M.L. (1987) Duckling responses to lake acidification. *Transaction North American Wildlife and Natural Resources Conference*, **52**, 636-644.
Dillon,P.J., Yan,N.D., Harvey,H.H. (1984) Acidic Deposition - Effects on Aquatic Ecosystems. *Crc Critical Reviews in Environmental Control*, **13**, 167-194.
Dixit,A.S., Dixit,S.S., Smol,J.P. (1992a) Long-term trends in lake water pH and metal concentrations inferred from diatoms and chrysophytes in 3 lakes near Sudbury, Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 17-24.
Dixit,S.S., Cumming,B.F., Birks,H.J.B., Smol,J.P., Kingston,J.C., Uutala,A.J., Charles,D.F., Camburn,K.E. (1993) Diatom assemblages from Adirondack lakes (New York, USA) and the development of inference models for retrospective environmental assessment. *Journal of Paleolimnology*, **8**, 27-47.
Dixit,S.S., Dixit,A.S., Smol,J.P. (1989) Lake acidification recovery can be monitored using chrysophycean microfossils. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 1309-1312.
Dixit,S.S., Dixit,A.S., Smol,J.P. (1992b) Assessment of changes in lake water chemistry in Sudbury area lakes since preindustrial times. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 8-16.
Dixit,S.S., Smol,J.P. (1994) Diatoms as indicators in the environmental monitoring and assessment program - surface waters (EMAP-SW). *Environmental Monitoring and Assessment*, **31**, 275-306.
Dodson,S.I., Frey,D.G. (1991) Cladocera and other Branchiopoda. In: *Ecology and Classification of North American Freshwater Invertebrates*, 723-786, Academic Press, Inc., Toronto.
Duigan,C.A. (1992) The ecology and distribution of the littoral freshwater Chydoridae (Branchiopoda, Anomopoda) of Ireland, with taxonomic comments on some species. *Hydrobiologia*, **241**, 1-70.
Duigan,C.A., Kovach,W.L. (1991) A study of the distribution and ecology of littoral freshwater chydorid (Crustacea, Cladocera) communities in Ireland using multivariate analyses. *Journal of Biogeography*, **18**, 267-280.

Eriksson,M.O.G. (1984) Acidification of lakes - effects on waterbirds in Sweden. *Ambio*, **13**, 260-262.
Evans,R.E., Brown,S.B., Hara,T.J. (1988) The effects of aluminum and acid on the gill morphology in Rainbow Trout, *Salmo gairdneri*. *Environmental Biology of Fishes*, **22**, 299-311.
Fauquette,S., Guiot,J., Suc,J.P. (1998) A method for climatic reconstruction of the Mediterranean Pliocene using pollen data. *Palaeogeography Palaeoclimatology Palaeoecology*, **144**, 183-201.
Flößner,D. (1972) *Krebstiere, Crustacea. Miemen- und Blattfüsser, Branchiopoda, Fischläuse, Brachiura.*, 1-501, G. Fischer, Jena.
Flower,R.J. (1986) The relationship between surface sediment diatom assemblages and pH in 33 Galloway lakes - some regression-models for reconstructing pH and their application to sediment cores. *Hydrobiologia*, **143**, 93-103.
Flower,R.J. (1993) Diatom preservation - experiments and observations on dissolution and breakage in modern and fossil material. *Hydrobiologia*, **269**, 473-484.
Flower,R.J., Battarbee,R.W. (1983) Diatom evidence for recent acidification of 2 Scottish lochs. *Nature*, **305**, 130-133.
Flower,R.J., Battarbee,R.W., Appleby,P.G. (1987) The recent paleolimnology of acid lakes in Galloway, Southwest Scotland - diatom analysis, pH trends, and the role of afforestation. *Journal of Ecology*, **75**, 797-824.
Flower,R.J., Battarbee,R.W., Stevenson,A.C., Patrick,S.T., Appleby,P.G., Beebee,T.J.C., Fletcher,C., Marsh,C., Natkanski,J. (1988) *A Palaeoecological evaluation of the acidification of Cranmer Pond, Hampshire*, 1-76, ENSIS Ltd., London.
Flower,R.J., Cameron,N.G., Rose,N.L., Fritz,S.C., Harriman,R., Stevenson,A.C. (1990) Post-1970 water-chemistry changes and paleolimnology of several acidified upland lakes in the UK. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 427-433.
Flower, R. J., Jones, V. J., Battarbee, R. W., Appleby, P. G., Rippey, B., Rose, N. L., and Stevenson, A. C. The extent of regional acidification in North-West Scotland: Palaeoecological evidences. ECRC Research Paper No. 8. 1993. London, Environmental Change Research Centre.
Flower,R.J., Juggins,S., Battarbee,R.W. (1997) Matching diatom assemblages in take sediment cores and modern surface sediment samples: The implications for lake conservation and restoration with special reference to acidified systems. *Hydrobiologia*, **344**, 27-40.
Flower,R.J., Rippey,B., Rose,N.L., Appleby,P.G., Battarbee,R.W. (1994) Paleolimnological evidence for the acidification and contamination of lakes by atmospheric-pollution in western Ireland. *Journal of Ecology*, **82**, 581-596.
Frey,D.G. (1959) The taxonomic and phylogenetic significance of the head pores of the genus *Chydoridae* (Cladocera). *Internationale Revue der Gesamten Hydrobiologie*, **44**, 27-50.
Frey,D.G. (1960a) On the occurrence of cladoceran remains in lake sediments. *Proceedings of the National Academy of Sciences of the United States of America*, **46**, 917-920.
Frey,D.G. (1960b) The ecological significance of cladoceran remains in lake sediments. *Ecology*, **41**, 684-699.
Frey,D.G. (1962a) Cladocera from the Eemian interglacial of Denmark. *Journal of Paleontology*, **36**, 1133-1155.
Frey,D.G. (1962b) Supplement to: The taxonomic and the phylogenetic significance of the head pores of the Chydoridae (Cladocera). *Internationale Revue der Gesamten Hydrobiologie*, **47**, 603-609.
Frey,D.G. (1964) Differentiation of *Alona costata* SARS from two related species (Cladocera, Chydoridae). *Crustaceana*, **8**, 159-173.
Frey,D.G. (1965) A new genus of Chydoridae (Cladocera). *Internationale Revue der Gesamten Hydrobiologie*, **50**, 153-168.
Frey,D.G. (1986) Cladocera analysis. In: *Handbook of Palaeoecology and Palaeohydrology*, 667-692, John Wiley and Sons., New York.
Frey,D.G. (1988) Littoral and offshore communities of diatoms, cladocerans and dipterous larvae, and their interpretation in paleolimnology. *Journal of Paleolimnology*, **1**, 179-191.
Fritz,S.C. (1990) 20th-Century salinity and water-level fluctuations in Devils Lake, North Dakota - Test of a diatom-based transfer function. *Limnology and Oceanography*, **35**, 1771-1781.
Fritz,S.C., Kreiser,A.M., Appleby,P.G., Battarbee,R.W. (1990) Recent acidification of upland lakes in North Wales: Palaeolimnological evidence. In: *Acid Waters in Wales*, 27-37, Kluwer Academic Publishers, Dordrecht.
Fryer,G. (1968) Evolution and adaptive radiation in the Chydoridae (Crustacea: Cladocera): a study in comparative functional morphology and ecology. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **254**, 221-385.
Fryer,G. (1980) Acidity and species diversity in freshwater crustacean faunas. *Freshwater Biology*, **10**, 41-45.
Fryer,G. (1985) Crustacean diversity in relation to the size of water bodies: some facts and problems. *Freshwater Biology*, **15**, 347-361.
Fryer,G. (1993) *The freshwater Crustacea of Yorkshire: a faunistic and ecological survey*, 1-312, Yorkshire Naturalists' Union & Leeds Philosophical and Literary Society.
Gauch,H.G., Jr. (1982) *Multivariate analysis in community ecology*, 1-288, Cambridge University Press, Cambridge.
Glew,J.R. (1991) Miniature gravity corer for recovering short sediment cores. *Journal of Paleolimnology*, **5**, 258-287.
Glew,J.R., Smol,J.P., Last,W.M. (2001) Sediment core collection and extrusion. In: *Tracking Environmental Change Using Lake Sediments: Volume 1: Basin analysis, coring, and chronological techniques*, 73-105, Kluwer Academic Publishers, Dordrecht.
Gorham,E. (1958) The influence and importance of daily weather conditions in the supply of chloride, sulphate and other ions to fresh waters from atmospheric precipitation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **241**, 147-178.
Goulden,C.E. (1964) The history of the Cladoceran fauna of Esthwaite water (England) and its limnological significance. *Archiv für Hydrobiologie*, **60**, 1-52.
Goulden,C.E., Frey,D.G. (1963) The Occurrence and Significance of Lateral Head Pores in the Genus *Bosmina* (Cladocera). *Internationale Revue der Gesamten Hydrobiologie*, **48**, 513-522.
Grahn,O. (1986) Vegetation Structure and Primary Production in Acidified Lakes in Southwestern Sweden. *Experientia*, **42**, 465-470.
Grahn,O., Hultberg,H., Kandner,L. (1974) Oligotrophication - A self-accelerating process in lakes subjected to excessive supply of acid substances. *Ambio*, **3**, 93-94.
Guiot,J., Pons,A., de Beaulieu,J.L., Reille,M. (1989) A 140,000-year continental climate reconstruction from 2 European pollen records. *Nature*, **338**, 309-313.
Gunn,J.M., Keller,W. (1990) Biological recovery of an acid lake after reductions in industrial emissions of sulfur. *Nature*, **345**, 431-433.
Hann,B.J., Turner,M.A. (2000) Littoral microcrustacea in Lake 302S in the Experimental Lakes Area of Canada: acidification and recovery. *Freshwater Biology*, **43**, 133-146.
Harmsworth,R.V. (1968) The developmental history of Blelham Tarn (England) as shown by animal microfossils, with special reference to the Cladocera. *Ecological Monographs*, **38**, 223-241.
Harriman,R., Morrison,B.R.S. (1982) Ecology of Streams Draining Forested and Non-Forested Catchments in An Area of Central Scotland Subject to Acid Precipitation. *Hydrobiologia*, **88**, 251-263.
Harriman,R., Morrison,B.R.S., Caines,L.A., Collen,P., Watt,A.W. (1987) Long-Term Changes in Fish Populations of Acid Streams and Lochs in Galloway South West Scotland. *Water, Air, and Soil Pollution*, **32**, 89-112.
Havas,M., Hutchinson,T.C., Likens,G.E. (1984) Effect of low pH on sodium regulation in two species of *Daphnia*. *Canadian Journal of Zoology*, **62**, 1965-1970.

Havas,M., Likens,G.E. (1985) Toxicity of Aluminum and Hydrogen Ions to *Daphnia catawba*, *Holopedium gibberum*, *Chaoborus punctipennis*, and *Chironomus anthrocinus* from Mirror Lake, New-Hampshire. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*, **63**, 1114-1119.

Havas,M., Rosseland,B.O. (1995) Response of zooplankton, benthos, and fish to acidification: An overview. *Water, Air, and Soil Pollution*, **85**, 51-62.

Havens,K.E., DeCosta,J. (1987) The role of aluminium contamination in determining phytoplankton and zooplankton responses to acidification. *Water, Air, and Soil Pollution*, **33**, 277-293.

Hellsten,M.E., Sundberg,P. (2000) Genetic variation in two sympatric European populations of *Bosmina* spp. (Cladocera) tested with RAPD markers. *Hydrobiologia*, **421**, 157-164.

Henriksen,A., Kamari,J., Wilander,A. (1992) Critical loads of acidity: Nordic surface waters. *Ambio*, **21**, 356-363.

Henriksen,A., Lien,L., Traaen,T.S., Rosseland,B.O., Sevaldrud,I.S. (1990) The 1000-lake survey in Norway 1986. In: *The Surface Waters Acidification Programme*, 199-213, Cambridge University Press, Cambridge.

Henrikson,L., Hindar,A., Thornelof,E. (1995) Freshwater liming. *Water, Air, and Soil Pollution*, **85**, 131-142.

Higgs,E.S. (1997) What is good ecological restoration? *Conservation Biology*, **11**, 338-348.

Hill,M.O. (1973) Reciprocal averaging: an Eigenvector method of ordination. *Journal of Ecology*, **61**, 237-249.

Hill,M.O., Gauch,H.G., Jr. (1980) Detrended correspondence analysis, an improved ordination technique. *Vegetatio*, **42**, 47-48.

Hofmann,W. (1984) Postglacial morphological variation in Bosmina longispina Leydig (Crustacea, Cladocera) from the Großer Plöner See (north Germany) and its taxonomic implications. *Zeitschrift für Zoologische Systematik und Evolutionsforschung*, **22**, 294-301.

Hofmann,W. (1986) Developmental history of the Grosser Plöner See and the Schöhsee (north Germany): cladoceran analysis, with special reference to eutrophication. *Archiv für Hydrobiologie*, **74**, 259-287.

Hofmann,W. (2000) Response of the chydorid faunas to rapid climatic changes in four alpine lakes at different altitudes. *Palaeogeography Palaeoclimatology Palaeoecology*, **159**, 281-292.

Hofmann,W. (2001) Late-Glacial/Holocene succession of the chironomid and cladoceran fauna of the Soppensee (Central Switzerland). *Journal of Paleolimnology*, **25**, 411-420.

Hornung,M., Newson,M.D. (1986) Upland afforestation: influences on stream hydrology and chemistry. *Soil Use Management*, **2**, 61-65.

Howells,G. (1995) *Acid rain and acid waters*, 1-262, Ellis Horwood Limited, Hemel Hemstead.

Howells,G., Dalziel,T.R.K., Turnpenny,A.W.H. (1992) Loch Fleet - Liming to Restore A Brown Trout Fishery. *Environmental Pollution*, **78**, 131-139.

Hultberg,H., Andersson,I.B. (1982) Liming of Acidified Lakes - Induced Long-Term Changes. *Water, Air, and Soil Pollution*, **18**, 311-331.

Ihaka,R., Gentleman,R. (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, **5**, 299-314.

Jackson,D.A. (1993) Stopping rules in principal components-analysis - a comparison of heuristic and statistical approaches. *Ecology*, **74**, 2204-2214.

Jackson,S.T., Charles,D.F. (1988) Aquatic Macrophytes in Adirondack (New-York) Lakes - Patterns of Species Composition in Relation to Environment. *Canadian Journal of Botany-Revue Canadienne de Botanique*, **66**, 1449-1460.

Jansson,M., Persson,G., Broberg,O. (1986) Phosphorus in Acidified Lakes - the Example of Lake Gardsjon, Sweden. *Hydrobiologia*, **139**, 81-96.

Jenkins,A., Whitehead,P.G., Cosby,B.J., Birks,H.J.B. (1990) Modelling long-term acidification - a comparison with diatom reconstructions and the implications for reversibility. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 435-440.

Jenkins,A., Ferrier,R.C., Cosby,B.J. (1997) A dynamic model for assessing the impact of coupled sulphur and nitrogen deposition scenarios on surface water acidification. *Journal of Hydrology*, **197**, 111-127.

Jones,V.J., Stevenson,A.C., Battarbee,R.W. (1986) Lake acidification and the land-use hypothesis - a mid-post- glacial analog. *Nature*, **322**, 157-158.

Jongman,R.H.G., ter Braak,C.J.F., Van Tongeren,O.F.R. (1995) *Data analysis in community and landscape ecology*, Cambridge University Press.

Kent,M., Coker,P. (1992) *Vegetation description and analysis: a practical approach*, John Wiley & Sons, Chichester.

Kingston,J.C., Birks,H.J.B. (1990) Dissolved organic carbon reconstructions from diatom assemblages in PIRLA Project lakes, North America. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 279-288.

Kingston,J.C., Birks,H.J.B., Uutala,A.J., Cumming,B.F., Smol,J.P. (1992) Assessing trends in fishery resources and lakes water aluminium from paleolimnological analyses of siliceous algae. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 116-127.

Kinniburgh, D. G. and Edmunds, W. M. The susceptibility of U.K. groundwaters to acid deposition. Hydrological Report, British Geological Survey 86/3. 1986. British Geological Survey.

Korhola,A. (1999) Distribution patterns of Cladocera in subarctic Fennoscandian lakes and their potential in environmental reconstruction. *Ecography*, **22**, 357-373.

Korhola,A., Olander,H., Blom,T. (2000) Cladoceran and chironomid assemblages as quantitative indicators of water depth in subarctic Fennoscandian lakes. *Journal of Paleolimnology*, **24**, 43-54.

Korhola,A., Rautio,M. (2001) Cladocera and other branchiopod crustaceans. In: *Tracking Environmental Change Using Lake Sediments: Volume 4: Zoological Indicators*, 5-41, Kluwer Academic Publishers, Dordrecht.

Korhola,A., Vasko,K., Toivonen,H.T.T., Olander,H. (2002) Holocene temperature changes in northern Fennoscandia reconstructed from chironomids using Bayesian modelling. *Quaternary Science Reviews*, **21**, 1841-1860.

Korinek,V. (1971) Comparative study of head pores in the genus *Bosmina* BAIRD (Crustacae, Cladocera). *Vestnik Ceskoslovenske Spolecnosti Zoologicke*, **35**, 275-296.

Korinek,V., Sacherova,V., Havel,L. (1997) Subgeneric differences in head shield and ephippia ultrastructure within the genus Bosmina Baird (Crustacea, Cladocera). *Hydrobiologia*, **360**, 13-23.

Korsman,T., Birks,H.J.B. (1996) Diatom-based water chemistry reconstructions from northern Sweden: A comparison of reconstruction techniques. *Journal of Paleolimnology*, **15**, 65-77.

Kotov,A.A. (1996) Morphology and postembryonic development of males and females of Bosmina longispina Leydig (Crustacea, Anomopoda) from a North Iceland population. *Hydrobiologia*, **341**, 187-196.

Krause-Dellin,D., Steinberg,C. (1984) Evidence of lake acidification by a novel biological pH-meter. *Environmental Technology Letters*, **5**, 403-406.

Kreiser,A.M., Appleby,P.G., Natkanski,J., Rippey,B., Battarbee,R.W. (1990) Afforestation and lake acidification - a comparison of 4 sites in Scotland. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 377-383.

Lacroix,G. (1989) Morphological and ecological differentiation within the *Bosmina longirostris sensu lato* complex (Cladocera). *Les Comptes rendus l'Academie des sciences*, **308**, 373-378.

Last,W.M., Smol,J.P. (2001a) *Tracking Environmental Change Using Lake Sediments: Volume 1: Basin Analysis, coring and chronological techniques*, Kluwer Academic Publishers, Dordrecht.

Last,W.M., Smol,J.P. (2001b) *Tracking Environmental Change Using Lake Sediments: Volume 2: Physical and Geochemical Methods*, Kluwer Academic Publishers, Dordrecht.

Leavitt,P.R., Vinebrooke,R.D., Donald,D.B., Smol,J.P., Schindler,D.W. (1997) Past ultraviolet radiation environments in lakes derived from fossil pigments. *Nature*, **388**, 457-459.
Legendre,P., Legendre,L. (1998) *Numerical Ecology*, 1-853, Elsevier Science B.V., Amsterdam.
Leino,R.L., Wilkinson,P., Anderson,J.G. (1987) Histopathological changes in the gills of Pearl Dace, *Semotilus margarita*, and Fathead Minnows, *Pimephales promleas*, from experimentally acidified Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 126-134.
Lieder,U. (1983a) Die arten der untergattung *Eubosmina* SELIGO, 1900 (Crustacea: Cladocera, Bosminidae). *Mitteilungen aus dem Zoologischen Museum in Berlin*, **59**, 195-292.
Lieder,U. (1983b) Revision of the Genus *Bosmina* BAIRD, 1845 (Crustacea, Cladocera). *Internationale Revue der Gesamten Hydrobiologie*, **68**, 121-139.
Lindberg,S.E., Garten,C.T.Jr. (1988) Sources of sulphur in forest canopy throughfall. *Nature*, **336**, 148-151.
Lindenschmidt,K.E., Hamblin,P.F. (1997) Hypolimnetic aeration in Lake Tegel, Berlin. *Water Research*, **31**, 1619-1628.
Little,J.L., Hall,R.I., Quinlan,R., Smol,J.P. (2000) Past trophic status and hypolimnetic anoxia during eutrophicaton and remediation of Gravenhurst Bay, Ontario: comparison of diatoms, chironomids, and historical records. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 333-341.
Locke,A. (1992) Factors influencing community structure along stress gradients: zooplankton responses to acidification. *Ecology*, **73**, 903-909.
Locke,A., Sprules,W.G. (1994) Effects of lake acidification and recovery on the stability of zooplankton food webs. *Ecology*, **75**, 498-506.
Lotter,A.F., Birks,H.J.B. (1997) The separation of the influence of nutrients and climate on the varve time-series of Baldeggersee, Switzerland. *Aquatic Sciences*, **59**, 362-375.
Lotter,A.F., Birks,H.J.B., Hofmann,W., Marchetto,A. (1997) Modern diatom, cladocera, chironomid, and chrysophyte cyst assemblages as quantitative indicators for the reconstruction of past environmental conditions in the Alps .I. Climate. *Journal of Paleolimnology*, **18**, 395-420.
Madsen,T.V., Olesen,B., Bagger,J. (2002) Carbon acquisition and carbon dynamics by aquatic isoetids. *Aquatic Botany*, **73**, 351-371.
Matuszek,J.E., Beggs,G.L. (1988) Fish Species Richness in Relation to Lake Area, pH, and Other Abiotic Factors in Ontario Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**, 1931-1941.
McAuley,D.G., Longcore,J.R. (1988) Survival of juvenile Ring Necked ducks on wetlands of different pH. *Journal of Wildlife Management*, **52**, 169-176.
McCullagh,P., Nelder,J.A. (1983) *Generalized Linear Models*, 1-261, Chapman and Hall.
McWilliams,P.G., Potts,W.T.W. (1978) The effects of pH and calcium concentrations on gill potentials in the brown trout, *Salmo trutta*. *Journal of Comparative Physiology*, **126**, 277-286.
Meriläinen,J. (1967) The diatom flora and the hydrogen-ion concentration of water. *Annales Botanici Fennici*, **4**, 51-58.
Mills,K.H., Chalanchuk,S.M., Mohr,L.C., Davies,I.J. (1987) Responses of Fish Populations in Lake-223 to 8 Years of Experimental Acidification. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 114-125.
Minshall,G.W., Minshall,J.N. (1978) Further evidence on the role of chemical factors in determining the distribution of benthic invertebrates in the River Duddon. *Archiv für Hydrobiologie*, **83**, 324-355.
Monteith,D.T., Evans,C.D. (2000) *UK Acid Waters Monitoring Network: 10 Year Report*, ENSIS Publishing, London.
Moss,B., Stansfield,J., Irvine,K., Perrow,M.R., Phillips,G. (1996) Progressive restoration of a shallow lake: A 12-year experiment in isolation, sediment removal and biomanipulation. *Journal of Applied Ecology*, **33**, 71-86.
Muniz,I.P. (1991) Fresh-Water Acidification - Its Effects on Species and Communities of Fresh-Water Microbes, Plants and Animals. *Proceedings of the Royal Society of Edinburgh Section B- Biological Sciences*, **97**, 227-254.
Muniz,I.P., Walløe,L. (1990) The influence of water quality and catchment characteristics on the survival of fish populations. In: *The Surface Waters Acidification Programme*, 327-339, Cambridge University Press, Cambridge.
NEGTAP. Transboundary Air Pollution: Acidification, Eutrophication and Ground-Level Ozone in the UK. i-314. 2001. DEFRA.
Nilssen,J.P., Sandøy,S. (1986) Acidification history and crustacean remains: some ecological obstacles. *Hydrobiologia*, **143**, 349-354.
Nilssen,J.P., Sandøy,S. (1990) Recent lake acidification and cladoceran dynamics - surface sediment and core analyses from lakes in Norway, Scotland and Sweden. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 299-309.
Nilssen,J.P., Wærvågen,S.B. (2002) Intensive fish predation: an obstacle to biological recovery following liming of acidified lakes? *Journal of Aquatic Ecosystem Stress and Recovery*, **9**, 73-84.
Nilsson,S.I. (1993) Acidification of Swedish oligotrophic lakes - interactions between deposition, forest growth and effects on lake-water quality. *Ambio*, **22**, 272-276.
Nilsson,S.I., Miller,H.G., Miller,J.D. (1982) Forest growth as a possible cause of soil and water acidification. *Oikos*, **39**, 40-49.
Odén, S. The acidification of air precipitation and its consequences in the natural environment. Energy Committee Bulletin 1 . 1968. Stockholm, Swedish Natural Sciences Research Council.
Økland,J. (1983) Factors Regulating the Distribution of Fresh-Water Snails (Gastropoda) in Norway. *Malacologia*, **24**, 277-288.
Økland,J., Økland,K.A. (1986) The Effects of Acid Deposition on Benthic Animals in Lakes and Streams. *Experientia*, **42**, 471-486.
Økland,R.H. (1999) On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, **10**, 131-136.
Olander,H., Birks,H.J.B., Korhola,A., Blom,T. (1999) An expanded calibration model for inferring lakewater and air temperatures from fossil chironomid assemblages in northern Fennoscandia. *Holocene*, **9**, 279-294.
Ormerod,S.J., Allinson,N., Hudson,D., Tyler,S.J. (1986) The distribution of breeding dippers (*Cinclus cinclus* (L), Aves) in relation to stream acidity in upland Wales. *Freshwater Biology*, **16**, 501-507.
Ormerod,S.J., O'Halloran,J., Gribbin,S.D., Tyler,S.J. (1991) The ecology of dippers *Cinclus cinclus* in relation to stream acidity in upland Wales - Breeding performance, calcium physiology and nestling growth. *Journal of Applied Ecology*, **28**, 419-433.
Overpeck,J.T., Webb,T., Prentice,I.C. (1985) Quantitative interpretation of fossil pollen spectra - dissimilarity coefficients and the method of modern analogs. *Quaternary Research*, **23**, 87-108.
Paterson,M.J. (1994) Paleolimnological reconstruction of recent changes in assemblages of Cladocera from acidified lakes in the Adirondack Mountains (New York). *Journal of Paleolimnology*, **11**, 189-200.
Patrick,S.T., Battarbee,R.W., Jenkins,A. (1996) Monitoring acid waters in the UK: An overview of the UK acid waters monitoring network and summary of the first interpretative exercise. *Freshwater Biology*, **36**, 131-150.
Patrick,S.T., Monteith,D.T., Jenkins,A. (1995) *UK Acid Waters Monitoring Network: the first five years. Analysis and interpretation of results, April 1988 - March 1993*, 1-320, ENSIS Publishing, London.
Pennington,W. (1984) Long-term natural acidification of upland sites in Cumbria: evidence from post glacial lake sediments. *Freshwater Biological Association Annual Report*, **52**, 28-46.
Peyron,O., Guiot,J., Cheddadi,R., Tarasov,P., Reille,M., de Beaulieu,J.L., Bottema,S., Andrieu,V. (1998) Climatic reconstruction in Europe for 18,000 yr B.P. from pollen data. *Quaternary Research*, **49**, 183-196.
Post,D.M., Frost,T.M., Kitchell,J.F. (1995) Morphological responses by *Bosmina longirostris* and *Eubosmina tubicen* to changes in copepod predator populations during a whole-lake acidification experiment. *Journal of Plankton Research*, **17**, 1621-1632.

Potts,W.T.W., Fryer,G. (1979) The effects of pH and salt content on the sodium balance in *Daphnia magna* and *Acantholeberis curvirostris*. *Journal of Comparative Physiology*, **129**, 289-94.
Pough,F.H. (1976) Acid precipitation and embryonic mortality of spotted salamanders, *Ambystoma maculatum*. *Science*, **192**, 68-70.
Pough,F.H., Wilson,R.E. (1977) Acid precipitation and reproductive success of *Ambystoma* salamanders. *Water, Air, and Soil Pollution*, **7**, 307-316.
Prentice,I.C. (1980) Multidimensional scaling as a research tool in Quaternary palynology: a review of theory and methods. *Review of Palaeobotany and Palynology*, **31**, 71-104.
Psenner,R., Schmidt,R. (1992) Climate-driven pH control of remote alpine lakes and effects of acid deposition. *Nature*, **356**, 781-783.
Quade,H.W. (1969) Cladoceran Faunas Associated with Aquatic Macrophytes in Some Lakes in Northwester Minnesota. *Ecology*, **50**, 170-179.
Rautio,M. (2001) Ecology of zooplankton in subarctic ponds, with a focus on responses to ultraviolet radiation. *Kilpisjärvi Notes*, **15**, 1-30.
Renberg,I. (1990) A 12600 year perspective of the acidification of Lilla-Oresjon, Southwest Sweden. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 357-361.
Renberg,I., Brodin,Y.M., Cronberg,G., Eldaoushy,F., Oldfield,F., Rippey,B., Sandøy,S., Wallin,J.E., Wik,M. (1990) Recent acidification and biological changes in Lilla-Oresjon, southwest Sweden, and the relation to atmospheric-pollution and land-use history. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 391-396.
Renberg,I., Hellberg,T. (1982) The pH History of Lakes in Southwestern Sweden, as Calculated from the Subfossil Diatom Flora of the Sediments. *Ambio*, **11**, 30-33.
Renberg,I., Hultberg,H. (1992) A paleolimnological assessment of acidification and liming effects on diatom assemblages in a Swedish lake. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 65-72.
Renberg,I., Korsman,T., Anderson,N.J. (1993a) A temporal perspective of lake acidification in Sweden. *Ambio*, **22**, 264-271.
Renberg,I., Korsman,T., Birks,H.J.B. (1993b) Prehistoric increases in the pH of acid-sensitive Swedish lakes caused by land-use changes. *Nature*, **362**, 824-827.
Ripley,B.D. (1994) Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society Series B*, **56**, 409-437.
Roberts,N. (1989) *The Holocene; an environmental history*, 1-227, Blackwell Publishers.
Roelfs,J.G.M. (1983) Impact of acidification and eutrophication on macrophyte communities in soft water lakes in the Netherlands. I, Field observations. *Aquatic Botany*, **17**, 139-158.
Romundstad,A.J., Sandøy,S. (1995) Liming of acidified rivers in Norway; An attempt to preserve and restore biological diversity in acidified regions. *Water, Air, and Soil Pollution*, **85**, 997-1002.
Rose,N.L. (1995) Carbonaceous particle record in lake-sediments from the arctic and other remote areas of the Northern-hemisphere. *The Science of the Total Environment*, **161**, 487-496.
Rose,N.L. (1996) Inorganic fly-ash spheres as pollution tracers. *Environmental Pollution*, **91**, 245-252.
Rose,N.L., Golding,P.N.E., Battarbee,R.W. (1996) Selective concentration and enumeration of tephra shards from lake sediment cores. *Holocene*, **6**, 243-246.
Rose,N.L., Harlock,S., Appleby,P.G., Battarbee,R.W. (1995) Dating of recent lake-sediments in the United-Kingdom and Ireland using spheroidal carbonaceous particle (SCP) concentration profiles. *Holocene*, **5**, 328-335.
Rose,N.L., Juggins,S., Watt,J., Battarbee,R.W. (1994) Fuel-type characterisation of spheroidal carbonaceous particles using surface chemistry. *Ambio*, **23**, 296-299.
Rose,N.L., Rippey,B. (2002) The historical record of PAH, PCB, trace metal and fly-ash particle deposition at a remote lake in north-west Scotland. *Environmental Pollution*, **117**, 121-132.
Rosenqvist,I.T. (1978) Alternative sources for acidification of river water in Norway. *The Science of the Total Environment*, **10**, 39-49.
Round,F.E. (1990) Diatom communities - their response to changes in acidity. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 243-249.
Sarmaja-Korjonen, K. Chydorid ephippia as indicators of environmental change - biostratigraphical evidence from two lakes in southern Finland. Holocene . 2002.
Schartau,A.K.L., Walseng,B., Snucins,E. (2001) Correlation between microcrustaceans and environmental variables along an acidification gradient in Sudbury, Canada. *Water, Air, and Soil Pollution*, **130**, 1325-1330.
Schindler,D.W., Curtis,P.J., Parker,B.R., Stainton,M.P. (1996) Consequences of climate warming and lake acidification for UV-B penetration in North American boreal lakes. *Nature*, **379**, 705-708.
Schindler,D.W., Mills,K.H., Malley,D.F., Findlay,D.L., Shearer,J.A., Davies,I.J., Turner,M.A., Linsey,G.A., Cruikshank,D.R. (1985) Long-term ecosystem stress - the effects of years of experimental acidification on a small lake. *Science*, **228**, 1395-1401.
Schofield,C.L., Driscoll,C.T. (1987) Fish Species Distribution in Relation to Water-Quality Gradients in the North Branch of the Moose River Basin. *Biogeochemistry*, **3**, 63-85.
Scott,D.W. (1979) On optimal and data-based histograms. *Biometrika*, **66**, 605-610.
Siegfried,C.A., Bloomfield,J.A., Sutherland,J.W. (1989) Acidity Status and Phytoplankton Species Richness, Standing Crop, and Community Composition in Adirondack, New-York, Usa Lakes. *Hydrobiologia*, **175**, 13-32.
Skjelkvale,B.L., Wright,R.F. (1998) Mountain lakes; Sensitivity to acid deposition and global climate change. *Ambio*, **27**, 280-286.
Smith,M.A. (1990) The ecophysiology of epilithic diatom communities of acid lakes in Galloway, Southwest Scotland. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **327**, 251-256.
Smol,J.P. (1990) Paleolimnology: recent advances and future challenges. *Memorie dell'Insituto Italiano di Idrobiologia*, **47**, 253-276.
Smol,J.P. (1995) Paleolimnological approaches to the evaluation and monitoring of ecosystem health: providing a history for environmental damage and recovery. In: *Evaluating and Monitoring the Health of Large-Scale Ecosystems*, 301-318, Springer-Verlag, Berlin Heidelberg.
Smol,J.P., Birks,H.J.B., Last,W.M. (2001a) *Tracking Environmental Change Using Lake Sediments: Volume 3: Terrestrial, algal and siliceous indicators*, Kluwer Academic Publishers, Dordrecht.
Smol,J.P., Birks,H.J.B., Last,W.M. (2001b) *Tracking Environmental Change Using Lake Sediments: Volume 4: Zoological Indicators*, Kluwer Academic Publishers, Dordrecht.
Smolders,A.J.P., Lucassen,E.C.H.E., Roelofs,J.G.M. (2002) The isoetid environment: biogeochemistry and threats. *Aquatic Botany*, **73**, 325-350.
Søndergaard,M., Jensen,J.P., Jeppesen,E., Møller,P.H. (2002) Seasonal dynamics in the concentrations and retention of phosphorus in shallow Danish lakes after reduced loading. *Aquatic Ecosystem Health and Management*, **5**, 19-29.
Søndergaard,M., Jeppesen,E., Jensen,J.P., Lauridsen,T. (2000) Lake restoration in Denmark. *Lakes and Reservoirs: Research and Management*, **5**, 151-159.
Stemberger,R.S., Lazorchek,J.M. (1994) Zooplankton assemblage responses to disturbance gradients. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 2435-2447.
Stenson,J.A.E., Eriksson,M.O.G. (1989) Ecological Mechanisms Important for the Biotic Changes in Acidified Lakes in Scandinavia. *Archives of Environmental Contamination and Toxicology*, **18**, 201-206.

Stevenson,A.C., Juggins,S., Birks,H.J.B., Anderson,D.S., Anderson,N.J., Battarbee,R.W., Berge,F., Davis,R.B., Flower,R.J., Haworth,E.Y., Jones,V.J., Kingston,J.C., Kreiser,A.M., Line,J.M., Munro,M.A.R., Renberg,I. (1991) *The Surface Waters Acidification Project Palaeolimnology Programme: modern diatom / lake-water chemistry data-set*, 1-86, ENSIS Publishing, London.

Stevenson,A.C., Patrick,S.T., Kreiser,A.M., Battarbee,R.W. (1987) *Palaeoecological evaluations of the recent acidification of susceptible lakes: methods utilised under DoE contract PECD 7/7/139 and the Royal Society SWAP project.*, 1-36, Palaeoecological Research Unit, University College London, London.

Stoner,J.H., Gee,A.S., Wade,K.R. (1984) The effects of acidification on the ecology of streams in the Upper Tywi catchment in West Wales. *Environmental Pollution Series A-Ecological and Biological*, **35**, 125-157.

Svenson,T., Dickson,W., Hellberg,J., Moberg,G., Munthe,N. (1995) The Swedish liming programme. *Water, Air, and Soil Pollution*, **85**, 1003-1008.

Szeroczynska,K. (1998) Palaeolimnological investigations in Poland based on Cladocera (Crustacea). *Palaeogeography Palaeoclimatology Palaeoecology*, **140**, 335-345.

ter Braak,C.J.F. (1986) Canonical correspondence-analysis - a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167-1179.

ter Braak, C. J. F. Unimodal models to relate species to environment. 1-152. 1987. University of Wageningen.

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

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

ter Braak,C.J.F., Juggins,S., Birks,H.J.B., Van der Voet,H. (1993) Weighted averaging partial least squares regression (WA-PLS): definition and comparison with other methods for species-environment calibration. In: *Multivariate Environmental Statistics*, 526-559, Elsevier Science Publishers.

ter Braak,C.J.F., Prentice,I.C. (1988) A theory of gradient analysis. *Advances in Ecological Research*, **18**, 271-317.

ter Braak,C.J.F., Smilauer,P. (2002) *CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (version 4.5)*, 1-500, Microcomputer Power, Ithaca, NY, USA.

Townsend,C.R., Hildrew,A.G., Francis,J. (1983) Community Structure in Some Southern English Streams - the Influence of Physicochemical Factors. *Freshwater Biology*, **13**, 521-544.

Tremel,B., Frey,S.E., Yan,N.D., Somers,K.M., Pawson,T.W. (2000) Habitat specificity of littoral Chydoridae (Crustacea, Branchiopoda, Anomopoda) in Plastic Lake, Ontario, Canada. *Hydrobiologia*, **432**, 195-205.

Ultsch,G.R., Gros,G. (1979) Mucus as a diffusion barrier to oxygen; Possible role in O2 uptake at low pH in carp (*Cyprinus carpio*) gills. *Comparative Biochemistry and Physiology*, **62A**, 685-689.

Underwood,J., Donald,A.P., Stoner,J.H. (1987) Investigations Into the Use of Limestone to Combat Acidification in 2 Lakes in West Wales. *Journal of Environmental Management*, **24**, 29-40.

Uutala,A.J. (1990) *Chaoborus* (Diptera: Chaoboridae) mandibles - paleolimnological indicators of the historical status of fish populations in acid-sensitive lakes. *Journal of Paleolimnology*, **4**, 139-151.

Vasko,K., Toivonen,T.T., Korhola,A. (2000) A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *Journal of Paleolimnology*, **24**, 243-250.

Venables,W.N., Ripley,B.D. (1999) *Modern Applied Statistics with S-PLUS*, 1-501, Spring-Verlag New York Inc., New York.

Vinebrooke,R.D., Dixit,S.S., Graham,M.D., Gunn,J.M., Chen,Y.W., Belzile,N. (2002) Whole-lake algal responses to a century of acidic industrial deposition on the Canadian Shield. *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 483-493.

Walker,I.R. (2001) Midges: Chironomidae and related Diptera. In: *Tracking Environmental Change Using Lake Sediments. Volume 4: Zoological Indicators*, 43-66, Kluwer Academic Publishers, Dordrecht.

Walker,I.R., Mott,R.J., Smol,J.P. (1991) Allerød-Younger Dryas lake temperatures from midge fossils in Atlantic Canada. *Science*, **253**, 1010-1012.

Walseng,B., Karlsen,L.R. (2001) Planktonic and littoral microcrustaceans as indices of recovery in limed lakes in SE Norway. *Water, Air, and Soil Pollution*, **130**, 1313-1318.

Walseng,B., Schartau,A.K.L. (2001) Crustacean communities in Canada and Norway: Comparison of species along a pH gradient. *Water, Air, and Soil Pollution*, **130**, 1319-1324.

Webb,T.I., Laseski,R.A., Bernabo,J.C. (1978) Sensing vegetational patterns with pollen data: choosing the data. *Ecology*, **59**, 1151-1163.

Wetzel,R.G. (1983) *Limnology*, Saunders College Publishing / Harcourt Brace College Publishers.

Whiteside,M.C. (1970) Danish Chydorid Cladocera: Modern ecology and core studies. *Ecological Monographs*, **40**, 79-118.

Whiteside,M.C., Williams,J.B., White,C.P. (1978) Seasonal Abundance and Pattern of Chydorid, Cladocera in Mud and Vegetative Habitats. *Ecology*, **59**, 1177-1188.

Whittaker,R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1-80.

Whittaker,R.H. (1967) Gradient analysis of vegetation. *Biological Review*, **42**, 207-264.

Wik,M., Renberg,I., Darley,J. (1986) Sedimentary records of carbonaceous particles from fossil-fuel combustion. *Hydrobiologia*, **143**, 387-394.

Wilson,S.E., Walker,I.R., Mott,R.J., Smol,J.P. (1993) Climatic and limnological changes associated with the Younger Dryas in Atlantic Canada. *Climate Dynamics*, **8**, 177-187.

Wright,H.E.Jr. (1967) The use of surface samples in Quaternary pollen analysis. *Review of Palaeobotany and Palynology*, **2**, 321-330.

Yan,N.D., Miller,G.E., Wile,I., Hitchin,G.G. (1985) Richness of aquatic macrophyte floras of soft-water lakes of differing pH and trace-metal content in Ontario, Canada. *Aquatic Botany*, **23**, 27-40.

Yan,ND., Stokes,P. (1978) Phytoplankton of an acid lake and its response to experimental alterations of pH. *Environmental Conservation*, **5**, 93-100.

Zimmer,D.J. (1987) Effects of low pH acclimation on cladocerans - clues to the interaction of physiology and ecology of acid lake zooplankton. *Annales de la Societe Royale Zoologique de Belgique*, **117**, 139-149.

Appendix A

Site Codes used in the UKAWDDS and the 83-lake diatom and Cladocera training set

Site Codes					
SiteCode	Name	Country	Grid	Easting	Northing
ACH	Loch na Achlaise	SCO	NN	310	480
ALWN	Llyn Alwen	CYM	SH	898	567
ARR	Loch Coire nan Arr	SCO	NG	808	422
ARTH	Loch Arthur	SCO	NX	904	688
BALA	Llyn Tegid	CYM	SH	905	347
BARE	Loch Barean	SCO	NX	861	557
BARL	Llyn Barlwyd	CYM	SH	713	486
BEIS	Loch na Beiste	SCO	NG	885	943
BER	Llyn Berwyn	CYM	SN	743	568
BHAI	Loch Coire a Bhaic	SCO	NC	247	295
BHAR	Loch Bharranch	SCO	NG	977	575
BODG	Llyn Bodgynedd	CYM	SH	762	593
BODL	Llyn Bodlyn	CYM	SH	648	238
BOGA	Loch nam Badan Boga	SCO	NH	99	930
BRAC	Loch nam Brac	SCO	NC	179	480
BREC	Lochenbreck	SCO	NX	643	655
BUGE	Llyn Bugeilyn	CYM	SN	822	923
BURNMT	Burnmoor Tarn	ENG	NY	184	44
BYCH	Llyn Bychan	CYM	SH	753	593
CADH	Loch Dubh Cadhafuaraich	SCO	NC	682	183
CE45A	Llynnoedd Ieuan	CYM	SN	794	812
CFEA	Loch na Claise Ferna	SCO	NC	201	468
CFYN	Llyn Cwm Ffynnion	CYM	SH	648	564
CHAM	Loch a Cham Alltain	SCO	NC	283	446
CHN	Loch Chon	SCO	NN	421	51
CLAI	Loch Clair	SCO	NG	999	574
CLON	Loch Clonyard	SCO	NX	857	554
CLYD	Llyn Clyd	CYM	SH	635	597
CNAM	Loch Coire nan Cnamh	SCO	NG	974	38
CON	Llyn Conwy	CYM	SH	780	463
COR	Loch Coire an Lochan	SCO	NH	943	4
CORN	Loch Bealach Cornaidh	SCO	NC	208	282
CRA	Loch Craggie	SCO	NC	625	75
CREI	Loch na Creige Duibhe	SCO	NC	5	118
CRIC	Loch na Cric	SCO	NC	166	37
CUAR	Loch na Cuaran	SCO	NC	292	238
CULF	Loch Cul Fraioch	SCO	NC	25	330
CUR	Loch na Curra	SCO	NG	823	800
CWBY	Llyn Cwm Bychan	CYM	SH	640	313
CWEL	Llyn Cwellyn	CYM	SH	560	549
CZSN51	Llyn Llech Owain	CYM	SN	569	151
CZSN66	Llyn Eiddwen	CYM	SN	606	670
CZSN90	Llyn Fach	CYM	SN	905	370
DALL	Loch Dallas	SCO	NJ	92	475
DCAL	Loch Dubh Camas an	SCO	NG	871	972
DEVOKE	Devoke Water	ENG	SD	163	970
DIWA	Llyn Diawaunedd	CYM	SH	685	536
DOI	Loch Doilet	SCO	NM	808	678
DOON	Loch Doon	SCO	NX	495	985
DUBH	Lochan an Dubha	SCO	NC	147	55

Site Codes					
SiteCode	Name	Country	Grid	Easting	Northing
DUH	Dubh Loch	SCO	NO	238	828
DUL	Llyn Dulyn	CYM	SH	662	244
EDNO	Llyn Edno	CYM	SH	663	497
ENO	Loch Enoch	SCO	NX	445	851
EUN	Loch nan Eun	SCO	NO	230	854
FAN	Llyn y Fan Fawr	CYM	SN	831	216
FEOI	Lochan Feoir	SCO	NC	229	252
FERN	Loch Fern	SCO	NX	863	624
FHI	Coire Fhionn Lochan	SCO	NR	902	459
FHIO	Lochan Fhionnlaidh	SCO	NC	191	103
FINL	Loch Finlas	SCO	NX	460	983
FLE	Loch Fleet	SCO	NX	560	697
FLEO	Loch Fleodach Coire	SCO	NC	275	248
FNOD	Llyn Fanod	CYM	SN	603	643
GAIN	Loch na Gaineimh	SCO	NC	765	304
GARN	Llyn y Garn	CYM	SH	762	377
GEIR	Llyn Geirionydd	CYM	SH	763	606
GLAN	Llyn Glanmerin	CYM	SN	755	991
GLAS	Llyn Glas	CYM	SH	601	547
GLFR	Llyn Glasfryn	CYM	SH	402	422
GLOY	Gloyw Llyn	CYM	SH	646	299
GLYN	Llyn Glaslyn	CYM	SN	826	641
GOD	Llyn Goddionduon	CYM	SH	754	585
GREENT	Greendale Tarn	ENG	NY	146	74
GRUA	Loch na Gruagaich	SCO	NC	243	158
GWYN	Llyn Gwynant	CYM	SH	644	516
GYN	Llyn Gynon	CYM	SN	800	647
HAIR	Loch na h-Airbhe	SCO	NH	103	924
HARR	Loch Harrow	SCO	NX	527	867
HHHH	Un Named H	SCO	NO	653	909
HIR	Llyn Hir	CYM	SN	789	675
HOID	Loch na h-Oidhche	SCO	NH	154	778
HOWI	Loch Howie	SCO	NX	697	834
HUID	Loch Bealach a Bhuinich	SCO	NC	264	256
IDWA	Llyn Idwal	CYM	SH	645	596
INVA	Lochinvar	SCO	NX	659	853
IOIG	Loch Ian Oig	SCO	NG	792	292
IRD	Llyn Irddyn	CYM	SH	630	220
KEMP	Loch Kemp	SCO	NH	612	323
KIRR	Loch Kierriereoch	SCO	NX	363	865
LACH	Lochan Lairig Cheile	SCO	NN	558	278
LAG	Llyn Llagi	CYM	SH	649	483
LAI	Loch Laidon	SCO	NN	380	542
LAMH	Loch a Mhadaidh	SCO	NH	199	732
LAR	Loch na Larach	SCO	NC	214	583
LCSL	Llyn Cwm Silyn Lower	CYM	SH	512	508
LCSU	Llyn Cwm Silyn Upper	CYM	SH	515	505
LDE	Loch Dee	SCO	NX	470	790
LENY	Llyn Llennych	CYM	SH	655	377
LGR	Loch Grannoch	SCO	NX	541	691
LLDU	Llyn Du	CYM	SH	564	425
LLGH	Long Loch of Glenhead	SCO	NX	446	808
LNEI	Loch Nan Eion	SCO	NG	925	508
LOCH	Loch Toll an Lochain	SCO	NH	74	832
LOD	Lochan Dubh	SCO	NM	895	710

Site Codes					
SiteCode	Name	Country	Grid	Easting	Northing
LOSG	Loch Bad an Losguiun	SCO	NH	158	38
LOWT	Low Tarn	ENG	NY	163	91
MABE	Loch Mayberry	SCO	NX	286	750
MACA	Loch Macatarick	SCO	NX	440	912
MANN	Loch Mannoch	SCO	NX	664	605
MEO4B	Llyn Cau	CYM	SH	716	125
MHIC	Loch Mich Leoid	SCO	NJ	8	347
MINN	Loch Minnoch	SCO	NX	530	857
MOAN	Loch Moan	SCO	NX	346	858
MUCK	Loch Muck	SCO	NS	513	7
MUIG	Loch Muigbhlaraidh	SCO	NH	635	830
MYMB	Llyn Mymbyr	CYM	SH	709	574
NABE	Loch na Beiste	SCO	NC	4	125
NAGA	Lochnagar	SCO	NO	252	859
NAHU	Loch Bealach na h-Uidhe	SCO	NC	264	256
NEUN	Loch na Eun	SCO	NC	232	298
NIGH	Lochan Nigheadh	SCO	NC	182	148
OCHI	Loch Ochiltree	SCO	NX	317	745
PARC	Llyn y Parc	CYM	SH	793	587
PENR	Llyn Penrhaiadr	CYM	SN	753	933
RIEC	Loch Riecawr	SCO	NX	434	934
RLGH	Round Loch of Glenhead	SCO	NX	450	805
RONA	Loch Ronald	SCO	NX	265	644
SAID	Loch coire na Saidhe	SCO	NC	450	360
SCOATT	Scoat Tarn	ENG	NY	159	104
SGAM	Loch Sgamhaiu	SCO	NH	100	530
SKAK	Loch Skae	SCO	NX	710	837
SKE	Loch Skerrow	SCO	NX	605	682
STRO	Loch Stroan	SCO	NX	644	704
TANN	Loch Tanna	SCO	NR	921	428
TARF	Loch Tarff	SCO	NH	425	100
TEAN	Loch Teanga	SCO	NF	818	383
TECW	Llyn Tecwyn	CYM	SH	629	370
TINK	Loch Tinker	SCO	NN	445	68
TOLL	Loch Tollaidh	SCO	NG	841	785
TROO	Loch Trool	SCO	NX	412	798
UAI	Lochan Uaine	SCO	NO	1	981
UIS	Loch Uisge	SCO	NM	808	550
UN02	Un-Named	SCO	NC	168	478
URR	Loch Urr	SCO	NX	760	845
VAL	Loch Valley	SCO	NX	445	817
WHIN	Loch Whinyeon	SCO	NX	625	608
WHIT	White Loch	SCO	NX	864	547
WOOD	Loch Woodhall	SCO	NX	673	675
YBI	Llyn y Bi	CYM	SH	670	265
YGAD	Llyn y Gadair	CYM	SH	648	564

Appendix B

Diatom species codes used in the thesis.

Diatom Species Codes	
TaxonCode	**TaxonName**
AC001A	Achnanthes lanceolata
AC002A	Achnanthes linearis
AC002B	Achnanthes linearis curta
AC004A	Achnanthes pseudoswazi
AC006A	Achnanthes clevei clevei
AC011A	Achnanthes peragalli
AC013A	Achnanthes minutissima minutissima
AC014A	Achnanthes austriaca austriaca
AC014B	Achnanthes austriaca minor
AC014C	Achnanthes austriaca helvetica
AC017A	Achnanthes kryophila kryophila
AC018A	Achnanthes laterostrata
AC019A	Achnanthes nodosa
AC022A	Achnanthes marginulata
AC023A	Achnanthes conspicua conspicua
AC024A	Achnanthes depressa
AC025A	Achnanthes flexella
AC025B	Achnanthes flexella alpestris
AC027A	Achnanthes holstii
AC028A	Achnanthes saxonica
AC029A	Achnanthes sublaevis
AC030A	Achnanthes umara
AC034A	Achnanthes suchlandtii
AC035A	Achnanthes pusilla pusilla
AC038A	Achnanthes lapponica
AC038B	Achnanthes lapponica fennica
AC039A	Achnanthes didyma didyma
AC042A	Achnanthes detha
AC043A	Achnanthes lapidosa
AC044A	Achnanthes levanderi
AC045A	Achnanthes bicapitata
AC046A	Achnanthes altaica
AC048A	Achnanthes scotica
AC083A	Achnanthes laevis
AC119A	Achnanthes saccula
AC136A	Achnanthes subatomoides
AC142A	Achnanthes kuelbsii
AC151A	Achnanthes abundans
AC161A	Achnanthes ventralis
AC167A	Achnanthes daonensis
AC9999	Achnanthes sp.
AM001B	Amphora ovalis pediculus
AM001C	Amphora ovalis libyca
AM001D	Amphora ovalis affinis
AM011A	Amphora libyca
AM012A	Amphora pediculus
AM9999	Amphora sp.
AP001A	Amphipleura pellucida
AS001A	Asterionella formosa formosa
AS003A	Asterionella ralfsii

Diatom Species Codes	
TaxonCode	TaxonName
AT009A	Actinocyclus roperi
AU001A	Aulacoseira italica italica
AU001C	Aulacoseira italica valida
AU002A	Aulacoseira ambigua
AU004A	Aulacoseira lirata lirata
AU004B	Aulacoseira lirata lacustris
AU004C	Aulacoseira lirata biseriata
AU004D	Aulacoseira lirata alpigena
AU005A	Aulacoseira distans distans
AU005D	Aulacoseira distans tenella
AU005E	Aulacoseira distans nivalis
AU005H	Aulacoseira distans alpigena
AU005J	Aulacoseira distans laevissima
AU009A	Aulacoseira islandica islandica
AU010A	Aulacoseira perglabra
AU010B	Aulacoseira perglabra floriniae
AU014A	Aulacoseira nygaardii
AU020A	Aulacoseira subarctica
AU023A	Aulacoseira tethera
AU031A	Aulacoseira alpigena
AU9999	Aulacoseira sp.
BR001A	Brachysira vitrea
BR001B	Brachysira vitrea lanceolata
BR003A	Brachysira serians
BR003B	Brachysira serians modesta
BR004A	Brachysira styriaca
BR006A	Brachysira brebissonii brebissonii
CA002A	Caloneis bacillum bacillum
CA003A	Caloneis silicula
CA005A	Caloneis bacillaris bacillaris
CA018A	Caloneis tenuis
CA9999	Caloneis sp.
CM001A	Cymbella ventricosa
CM003A	Cymbella sinuata sinuata
CM004A	Cymbella microcephala microcephala
CM009A	Cymbella naviculiformis
CM010A	Cymbella perpusilla
CM013A	Cymbella helvetica helvetica
CM014A	Cymbella aequalis
CM015A	Cymbella cesatii cesatii
CM015B	Cymbella cesatii capitata
CM016A	Cymbella amphicephala amphicephala
CM017A	Cymbella hebridica
CM018A	Cymbella gracilis
CM019A	Cymbella lacustris
CM020A	Cymbella gaeumannii
CM022A	Cymbella affinis
CM029A	Cymbella ehrenbergii
CM031A	Cymbella minuta minuta
CM031C	Cymbella minuta silesiaca
CM035A	Cymbella angustata
CM038A	Cymbella delicatula
CM043A	Cymbella naviculacea
CM047A	Cymbella incerta
CM048A	Cymbella lunata

Diatom Species Codes	
TaxonCode	TaxonName
CM049A	Cymbella failaisensis
CM050A	Cymbella subaequalis
CM051A	Cymbella elginensis
CM052A	Cymbella descripta
CM101A	Cymbella scotica naviculacea
CM103A	Cymbella silesiaca
CM9999	Cymbella sp.
CO001A	Cocconeis placentula placentula
CO001B	Cocconeis placentula euglypta
CO005A	Cocconeis pediculus
CY001A	Cyclotella comta comta
CY002A	Cyclotella pseudostelligera
CY003A	Cyclotella meneghiniana meneghiniana
CY004A	Cyclotella stelligera
CY006A	Cyclotella kuetzingiana kuetzingiana
CY006B	Cyclotella kuetzingiana planetophora
CY006D	Cyclotella kuetzingiana minor
CY007A	Cyclotella glomerata
CY010A	Cyclotella comensis
CY011A	Cyclotella atomus
CY019A	Cyclotella radiosa
CY022A	Cyclotella bodanica
CY052A	Cyclotella rossii
CY054A	Cyclotella krammeri
CY9999	Cyclotella sp.
DE001A	Denticula tenuis tenuis
DP001A	Diploneis ovalis
DP003A	Diploneis oculata
DP007A	Diploneis oblongella oblongella
DP9999	Diploneis sp.
DT001A	Diatoma elongatum
DT001D	Diatoma elongatum tenue
DT002A	Diatoma hyemale hyemale
DT002B	Diatoma hyemale mesodon
DT003A	Diatoma vulgare vulgare
DT004B	Diatoma tenue elongatum
DT9999	Diatoma sp.
EP9999	Epithemia sp.
EU002A	Eunotia pectinalis pectinalis
EU002B	Eunotia pectinalis minor
EU002C	Eunotia pectinalis ventralis
EU002D	Eunotia pectinalis undulata
EU002E	Eunotia pectinalis minor impressa
EU002K	Eunotia pectinalis ventricosa
EU003A	Eunotia praerupta praerupta
EU004A	Eunotia tenella
EU007A	Eunotia bidentula
EU008A	Eunotia monodon monodon
EU009A	Eunotia exigua exigua
EU009C	Eunotia exigua tridentula
EU011A	Eunotia rhomboidea
EU013A	Eunotia arcus arcus
EU014A	Eunotia bactriana
EU015A	Eunotia denticulata denticulata
EU016A	Eunotia diodon

Diatom Species Codes	
TaxonCode	TaxonName
EU017A	Eunotia flexuosa flexuosa
EU018A	Eunotia formica
EU019A	Eunotia iatriaensis
EU020A	Eunotia meisteri meisteri
EU021A	Eunotia sudetica
EU022A	Eunotia bigibba bigibba
EU025A	Eunotia fallax
EU026A	Eunotia praerupta-nana
EU027A	Eunotia trinacria trinacria
EU028A	Eunotia microcephala
EU028B	Eunotia microcephala tridentata
EU029A	Eunotia valida
EU031A	Eunotia septentrionalis septentrionalis
EU032A	Eunotia serra serra
EU034A	Eunotia parallela parallela
EU039A	Eunotia triodon
EU040A	Eunotia paludosa
EU043A	Eunotia elegans
EU044A	Eunotia acmocephala
EU045A	Eunotia nymanniana
EU047A	Eunotia incisa
EU048A	Eunotia naegelii
EU049A	Eunotia curvata curvata
EU049B	Eunotia curvata subarcuata
EU049D	Eunotia curvata attenuata
EU050B	Eunotia tibia bidens
EU051A	Eunotia vanheurckii vanheurckii
EU051B	Eunotia vanheurckii intermedia
EU052A	Eunotia nodosa
EU053A	Eunotia tridentula
EU053B	Eunotia tridentula perminuta
EU056A	Eunotia minutissima
EU057A	Eunotia exgracilis
EU058A	Eunotia schwabei
EU060A	Eunotia pirla
EU070A	Eunotia bilunaris
EU070B	Eunotia bilunaris mucophila
EU106A	Eunotia rhynchocephela
EU107A	Eunotia implicata
EU111A	Eunotia soleirolii
EU9999	Eunotia sp.
FR001A	Fragilaria pinnata pinnata
FR001B	Fragilaria pinnata lancettula
FR001D	Fragilaria pinnata trigona
FR002A	Fragilaria construens construens
FR002C	Fragilaria construens venter
FR002D	Fragilaria construens exigua
FR005A	Fragilaria virescens virescens
FR005D	Fragilaria virescens exigua
FR006A	Fragilaria brevistriata brevistriata
FR007A	Fragilaria vaucheriae vaucheriae
FR008A	Fragilaria crotonensis
FR009A	Fragilaria capucina capucina
FR009B	Fragilaria capucina mesolepta
FR009F	Fragilaria capucina lanceolata

Diatom Species Codes	
TaxonCode	TaxonName
FR009H	Fragilaria capucina gracilis
FR010A	Fragilaria constricta constricta
FR011A	Fragilaria lapponica
FR013A	Fragilaria oldenburgiana
FR015A	Fragilaria lata
FR018A	Fragilaria elliptica
FR019A	Fragilaria intermedia
FR045A	Fragilaria parasitica
FR9999	Fragilaria sp.
FU002A	Frustulia rhomboides rhomboides
FU002B	Frustulia rhomboides saxonica
FU002F	Frustulia rhomboides viridula
GO003A	Gomphonema angustatum angustatum
GO004A	Gomphonema gracile
GO006A	Gomphonema acuminatum acuminatum
GO006C	Gomphonema acuminatum coronatum
GO010A	Gomphonema constrictum
GO013A	Gomphonema parvulum parvulum
GO014A	Gomphonema intricatum
GO014B	Gomphonema intricatum pumilum
GO017A	Gomphonema lanceolatum
GO023A	Gomphonema truncatum truncatum
GO025B	Gomphonema vibrio intricatum
GO025F	Gomphonema vibrio pumilum
GO050A	Gomphonema minutum
GO9999	Gomphonema sp.
HN001A	Hannaea arcus arcus
KR001A	Krasskella kriegerana
ME017A	Melosira nivalis
ME019A	Melosira arentii
ME044A	Melosira teres
MR001A	Meridion circulare circulare
NA002A	Navicula jaernefeltii
NA003A	Navicula radiosa radiosa
NA003B	Navicula radiosa tenella
NA005A	Navicula seminulum
NA005B	Navicula seminulum intermedia
NA006A	Navicula mediocris
NA006B	Navicula mediocris atomus
NA007A	Navicula cryptocephala cryptocephala
NA008A	Navicula rhynchocephala rhyncocephala
NA013A	Navicula pseudoscutiformis
NA014A	Navicula pupula pupula
NA014D	Navicula pupula mutata
NA014F	Navicula pupula elliptica
NA015A	Navicula hassiaca
NA016A	Navicula indifferens
NA029A	Navicula gracilis
NA030A	Navicula menisculus menisculus
NA032A	Navicula cocconeiformis cocconeiformis
NA033A	Navicula subtilissima
NA036A	Navicula perpusilla
NA037A	Navicula angusta
NA038A	Navicula arvensis
NA039A	Navicula festiva

Diatom Species Codes	
TaxonCode	TaxonName
NA042A	Navicula minima minima
NA043A	Navicula subatomoides
NA044A	Navicula krasskei
NA045A	Navicula bryophila bryophila
NA046A	Navicula contenta contenta
NA048A	Navicula soehrensis soehrensis
NA057A	Navicula elginensis elginensis
NA063A	Navicula trivialis
NA066A	Navicula capitata capitata
NA066B	Navicula capitata hungarica
NA068A	Navicula impexa
NA075A	Navicula subhamulata
NA084A	Navicula atomus
NA086A	Navicula tantula
NA099A	Navicula bremensis
NA101A	Navicula jaagii
NA102A	Navicula laevissima
NA112D	Navicula minuscula muralis
NA113A	Navicula acceptata
NA114A	Navicula subrotundata
NA115A	Navicula difficillima
NA124A	Navicula molestiformis
NA129A	Navicula seminuloides
NA133A	Navicula schassmannii
NA135A	Navicula tenuicephala
NA140A	Navicula madumensis
NA149A	Navicula digitulus
NA151A	Navicula gysingensis
NA156A	Navicula leptostriata
NA158A	Navicula cumbriensis
NA158B	Navicula cumbriensis minor
NA159A	Navicula menda
NA160A	Navicula submolesta
NA167A	Navicula hoefleri
NA168A	Navicula vitabunda
NA177A	Navicula abaujensis
NA399A	Navicula globosa
NA512A	Navicula minusculoides
NA532A	Navicula nodosa
NA631D	Navicula serians thermalis
NA667A	Navicula subsolaris
NA738A	Navicula vitiosa
NA751A	Navicula cryptotenella
NA9999	Navicula sp.
NE003A	Neidium affine affine
NE003B	Neidium affine longiceps
NE003C	Neidium affine amphirhynchus
NE004A	Neidium bisulcatum bisulcatum
NE006A	Neidium alpinum
NE012A	Neidium glaberrimum
NE020A	Neidium hercynicum
NE036A	Neidium ampliatum
NE9999	Neidium sp.
NI002A	Nitzschia fonticola
NI005A	Nitzschia perminuta

Diatom Species Codes	
TaxonCode	TaxonName
NI008A	Nitzschia frustulum
NI008D	Nitzschia frustulum perminuta
NI009A	Nitzschia palea palea
NI014A	Nitzschia amphibia amphibia
NI015A	Nitzschia dissipata
NI017A	Nitzschia gracilis
NI020A	Nitzschia angustata angustata
NI020B	Nitzschia angustata acuta
NI025A	Nitzschia recta
NI026A	Nitzschia romana
NI027A	Nitzschia microcephala
NI152A	Nitzschia pusilla
NI202A	Nitzschia alpina
NI9999	Nitzschia sp.
OP001A	Opephora martyi
OP9999	Opephora sp.
PE002A	Peronia fibula
PI004A	Pinnularia interrupta
PI005A	Pinnularia major major
PI007A	Pinnularia viridis viridis
PI008A	Pinnularia divergens divergens
PI011A	Pinnularia microstauron microstauron
PI012A	Pinnularia borealis
PI014A	Pinnularia appendiculata
PI015A	Pinnularia abaujensis abaujensis
PI016A	Pinnularia divergentissima divergentissima
PI018A	Pinnularia biceps biceps
PI019A	Pinnularia legumen legumen
PI020A	Pinnularia undulata
PI022A	Pinnularia subcapitata subcapitata
PI022B	Pinnularia subcapitata hilseana
PI023A	Pinnularia irrorata
PI030A	Pinnularia acoricola
PI040A	Pinnularia polyonca
PI042A	Pinnularia nodosa nodosa
PI9999	Pinnularia sp.
SA001A	Stauroneis anceps anceps
SA001B	Stauroneis anceps gracilis
SA004A	Stauroneis alpina
SA005A	Stauroneis legumen
SA006A	Stauroneis phoenicenteron phoenicenteron
SA9999	Stauroneis sp.
SE001A	Semiorbis hemicyclus
SP001A	Stenopterobia intermedia
SP002A	Stenopterobia sigmatella
ST004A	Stephanodiscus minutula
ST010A	Stephanodiscus parvus
SU004A	Surirella biseriata biseriata
SU005A	Surirella linearis linearis
SU006A	Surirella delicatissima delicatissima
SU010A	Surirella robusta robusta
SU016A	Surirella minuta
SU9999	Surirella sp.
SY001A	Synedra ulna ulna
SY002A	Synedra rumpens rumpens

Diatom Species Codes	
TaxonCode	TaxonName
SY003A	Synedra acus acus
SY004A	Synedra parasitica parasitica
SY008A	Synedra pulchella pulchella
SY009A	Synedra nana
SY010A	Synedra minuscula
SY013A	Synedra tenera
SY9999	Synedra sp.
TA001A	Tabellaria flocculosa flocculosa
TA001B	Tabellaria flocculosa flocculosa IIIp
TA001D	Tabellaria flocculosa III
TA002A	Tabellaria fenestrata
TA003A	Tabellaria binalis
TA004A	Tabellaria quadriseptata
TA9996	Tabellaria flocculosa agg.
TA9999	Tabellaria sp.
TE001A	Tetracyclus lacustris

Cladoceran species codes used in the thesis.

Cladocera Species Codes	
Taxon Code	Taxon Name
Acr harp	Acroperus harpae
Aln dent	Alonella dentifera
Aln exci	Alonella excisa
Aln exig	Alonella exigua
Aln glob	Alonella globulosa
Aln nana	Alonella nana
Aln rost	Alonella rostrata
Alo affi	Alona affinis
Alo cost	Alona costata
Alo gtub	Alona guttata tuberculata
Alo gutt	Alona guttata
Alo inte	Alona intermedia
Alo quad	Alona quadrangularis
Alo rect	Alona rectangula
Alo rust	Alona rustica
Alp elon	Acroperus elongata
Anc emar	Anchistropus emarginatus
Bos core	Bosmina coregoni
Bos lonr	Bosmina longirostris
Bos lons	Bosmina longispina
Bos spp	Bosmina Spp.
Cam rec	Camptocercus rectirostris
Cer spp	Ceriodaphnia Spp.
Chy gibb	Chydorus gibbus
Chy oval	Chydorus ovalis
Chy pige	Chydorus piger
Chy spha	Chydorus sphaericus
Dap long	Daphnia longispina group
Dap pule	Daphnia pulex group
Dap spp	Daphnia Spp. Sub Daphnia
Dia brac	Diaphanosoma brachyurum
Eur lame	Eurycercus lamellatus

Cladocera Species Codes	
Taxon Code	**Taxon Name**
Gra test	Graptoleberis testudinaria
Ily sord	Ilyocryptus c.f. sordidus
Lar alon	Large Alon spp.
Lat seti	Latona setifera
Ley acan	Leydigia acanthocercoides
Ley leyd	Leydigia leydigi
Mon disp	Monospilus dispar
Oxy tenu	Oxyurella tenuicaudis
Per trun	Percantha truncata
Ple trig	Pleuroxus trigonellus
Ple unci	Pleuroxus uncinatus
Pse glob	Pseudochydorus globosus
Rhy falc	Rhychotalona falcata
Sid crys	Sida crystallina
Sim spp	Simocephalus Spp.
Sma alon	Small Alona spp.
Unk nown	Unknown spp

Appendix C

Hydrochemical variable codes used in the thesis.

Hydrochemical Variables		
Code	**Name**	**Units**
Alk1	Alkalinity1	micro eq / l
Allab	Aluminium (labile)	micro g / l
Almon	Aluminium (monomeric)	micro g / l
Altot	Aluminium (Total)	micro g / l
Ca	Calcium	micro eq / l
Cl	Chloride	micro eq / l
Cond	Conductivity	u S / l
EqAlk	Equivalent Alkalinity	micro eq / l
K	Potassium	micro eq / l
Mg	Magnesium	micro eq / l
Na	Sodium	micro eq / l
NO3	Nitrate	micro eq / l
pH	pH	H+
SO4	Sulphate	micro eq / l
TOC	Total Organic Carbon	mg / l

Physical variable codes used in the thesis.

Physical Variables		
Code	**Name**	**Units**
Afforest	% Afforestation	%
Area:Dep	Ratio of Lake Area to Lake Depth	Ratio
CA:LA	Ratio of Catchment Area to Lake Area	Ratio
CArea	Catchment Area	ha
LAlt	Lake Altitude	m
LArea	Lake Area	ha
MaxAlt	Max Altitude in Catchment	m
MLDepth	Max Lake Depth	m
NCR	Net Catchment Relief	m

Appendix D

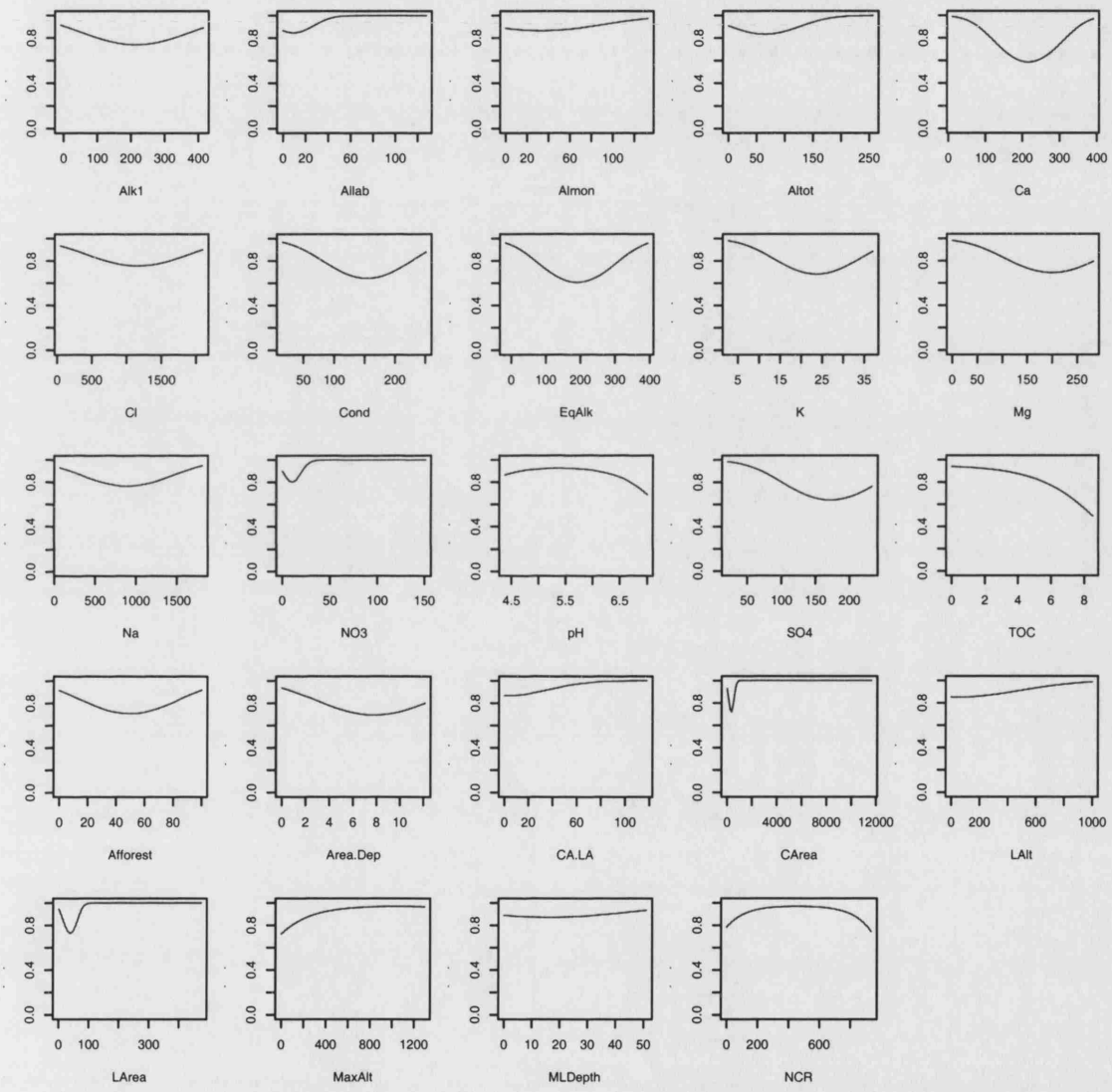


Figure 90: Generalised linear models of the response of Acroperus harpae to physicochemical parameters. The y-axis is the modelled probability of occurrence.

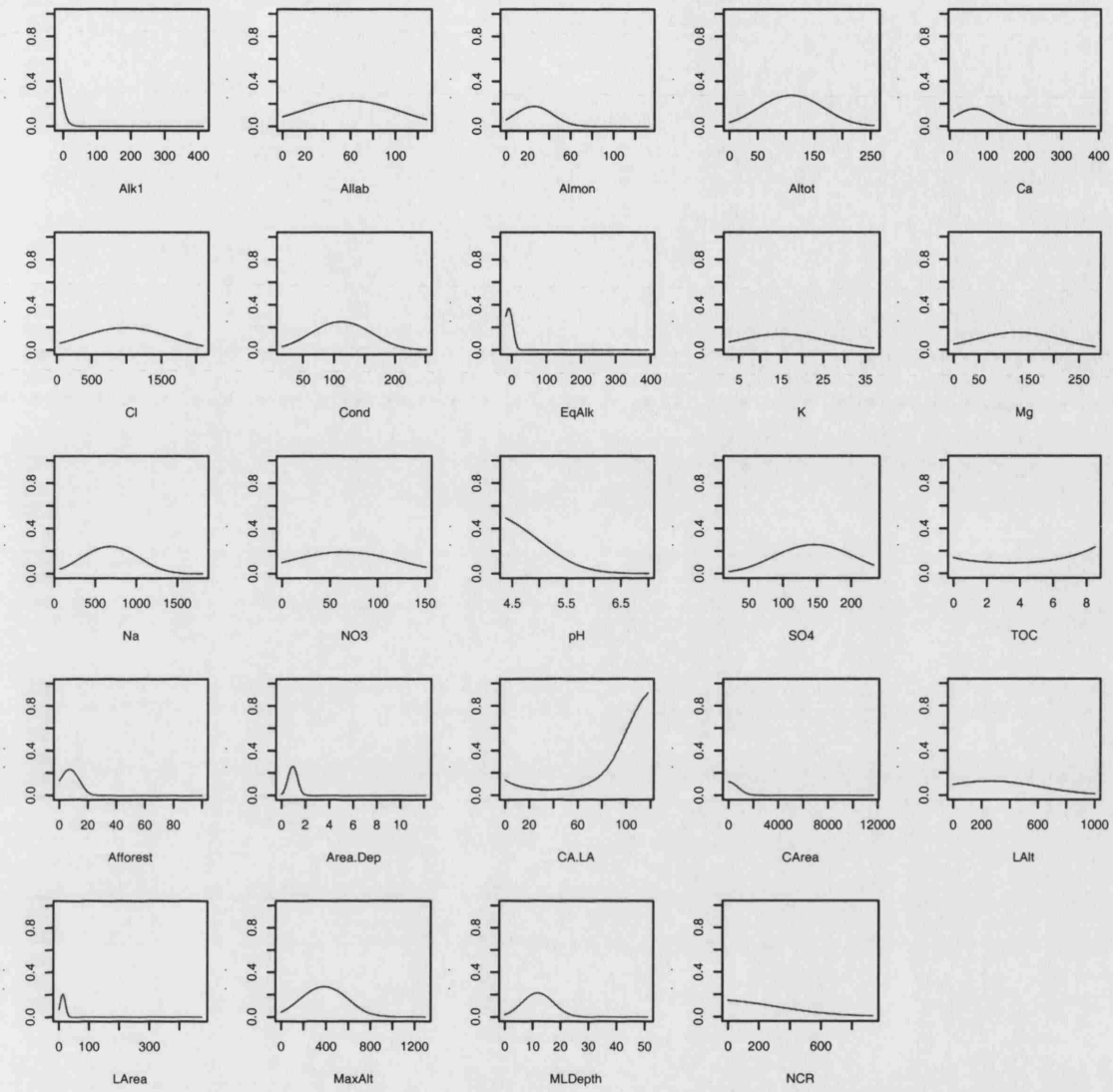


Figure 91: Generalised linear models of the response of Alonella dentifera to physicochemical parameters. The y-axis is the modelled probability of occurrence.

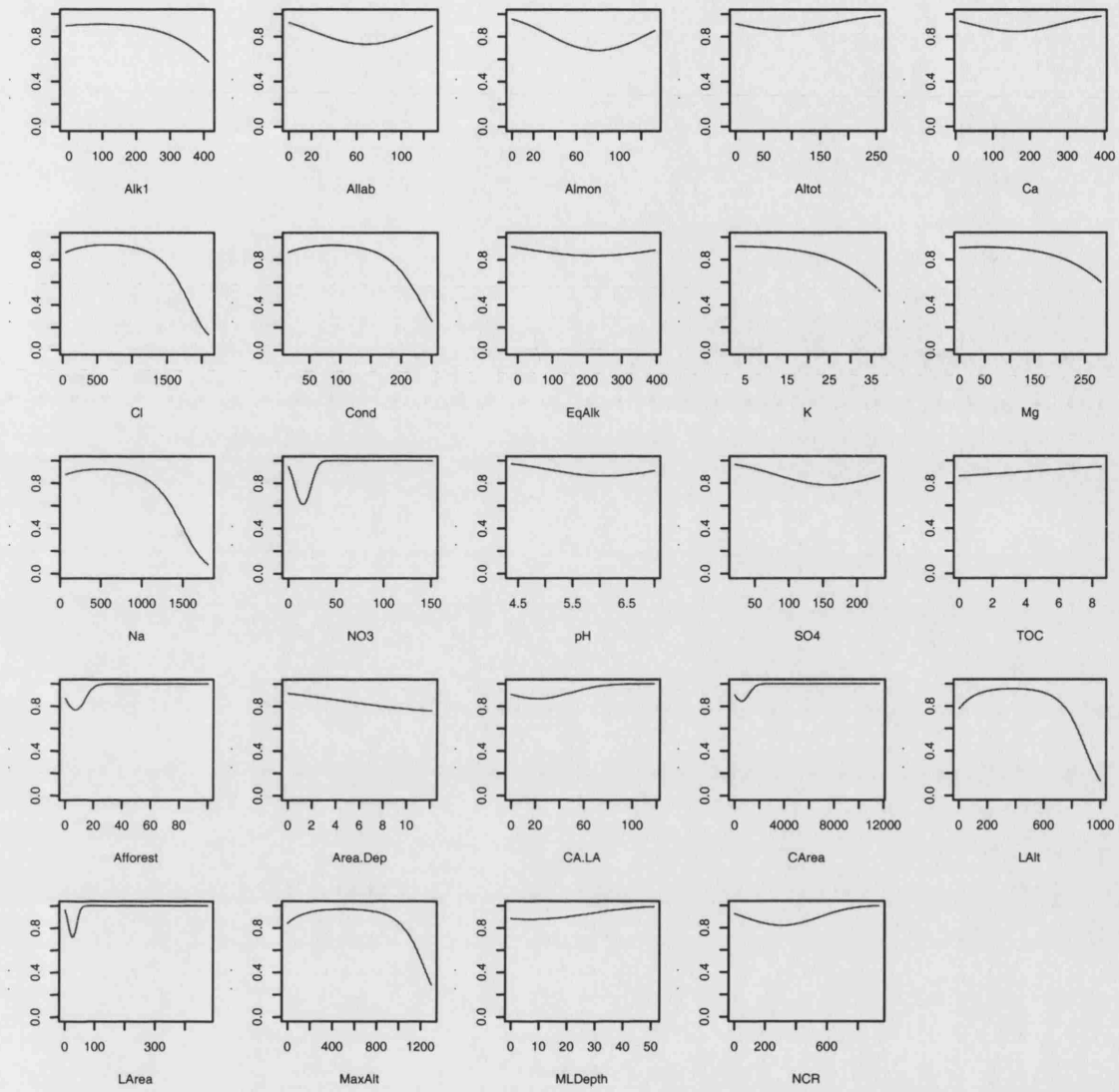


Figure 92: Generalised linear models of the response of Alonella excisa to physicochemical parameters. The y-axis is the modelled probability of occurrence.

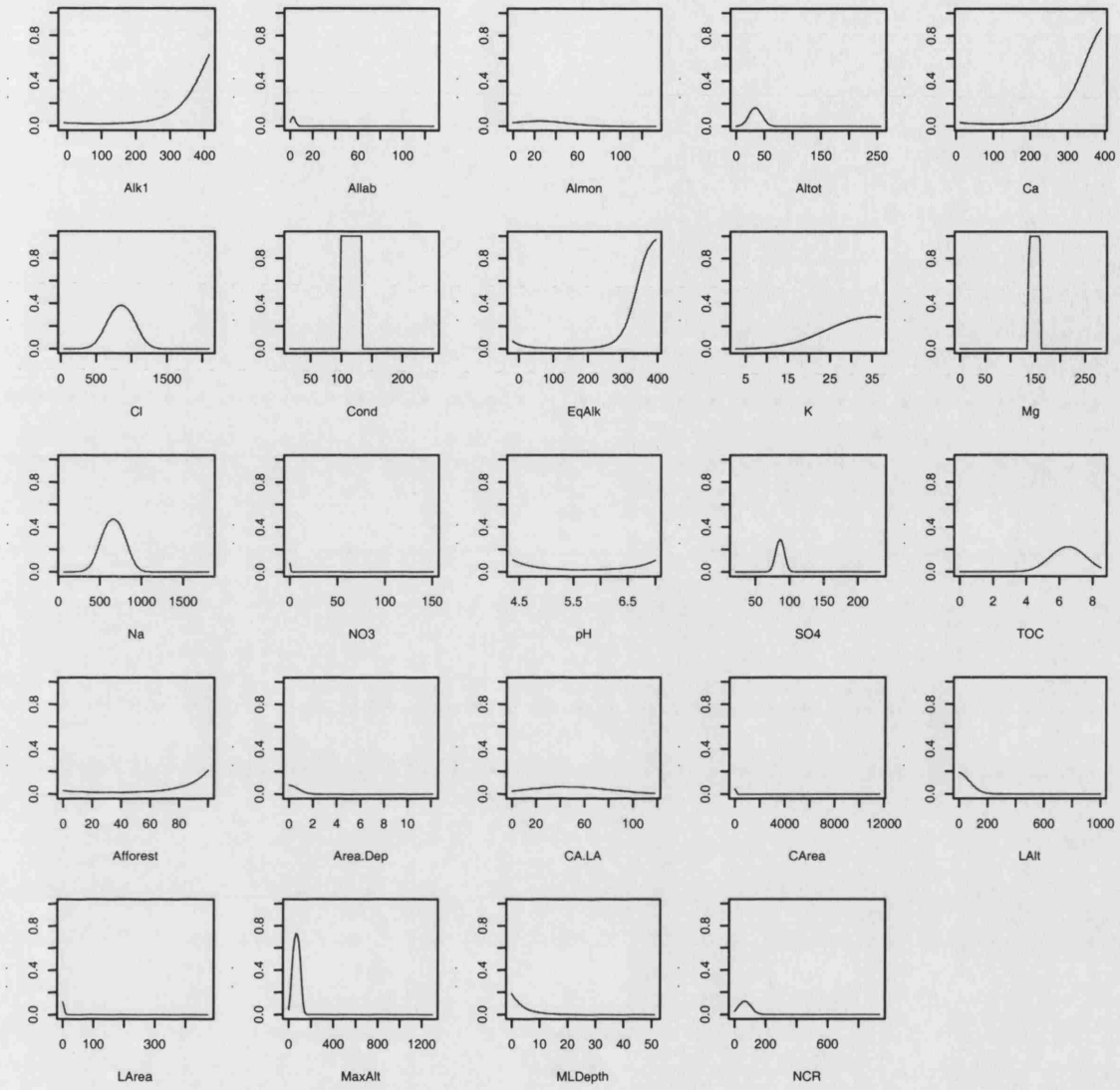


Figure 93: Generalised linear models of the response of Alonella exigua to physicochemical parameters. The y-axis is the modelled probability of occurrence.

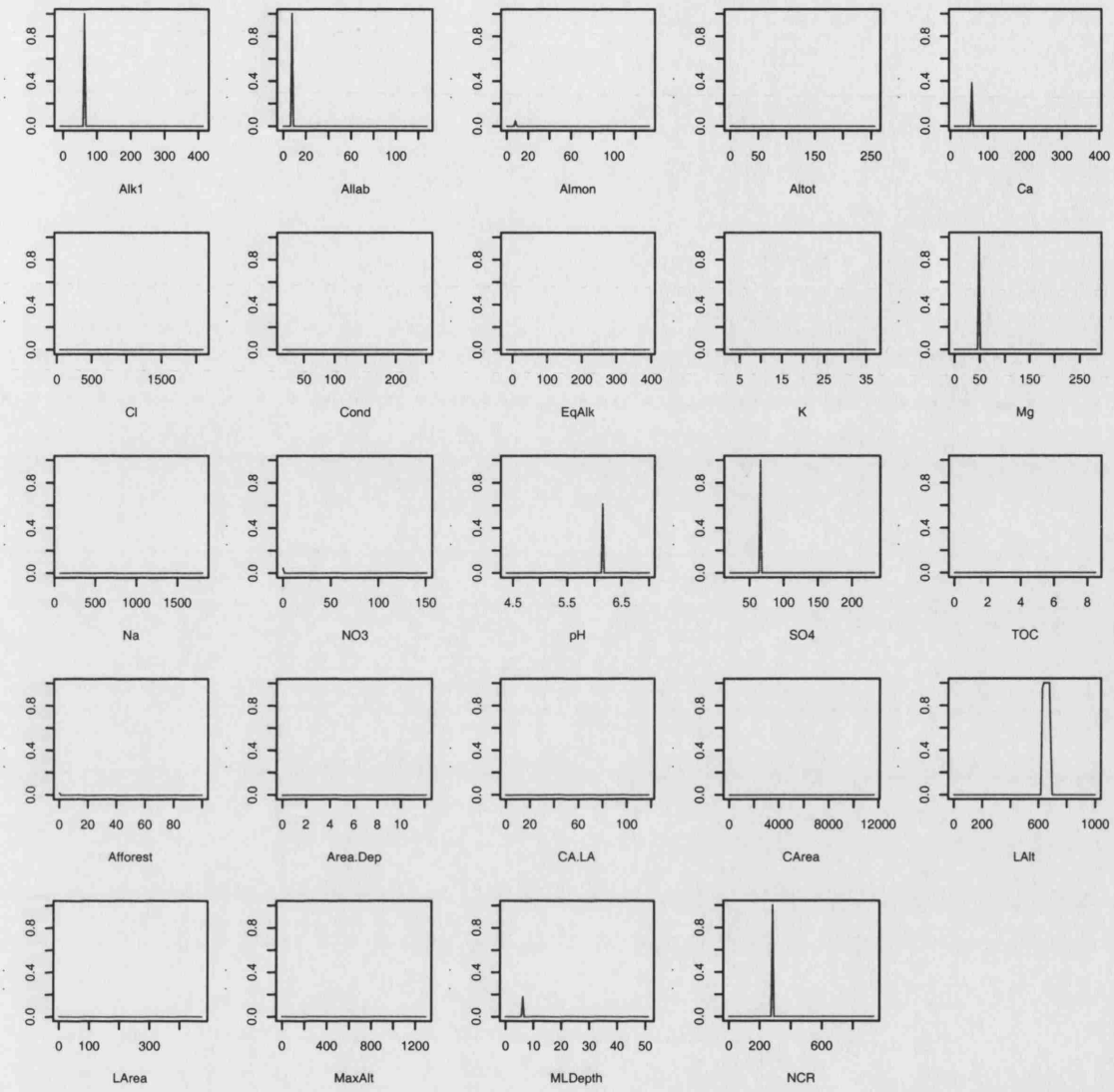


Figure 94: Generalised linear models of the response of Alonella globosus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

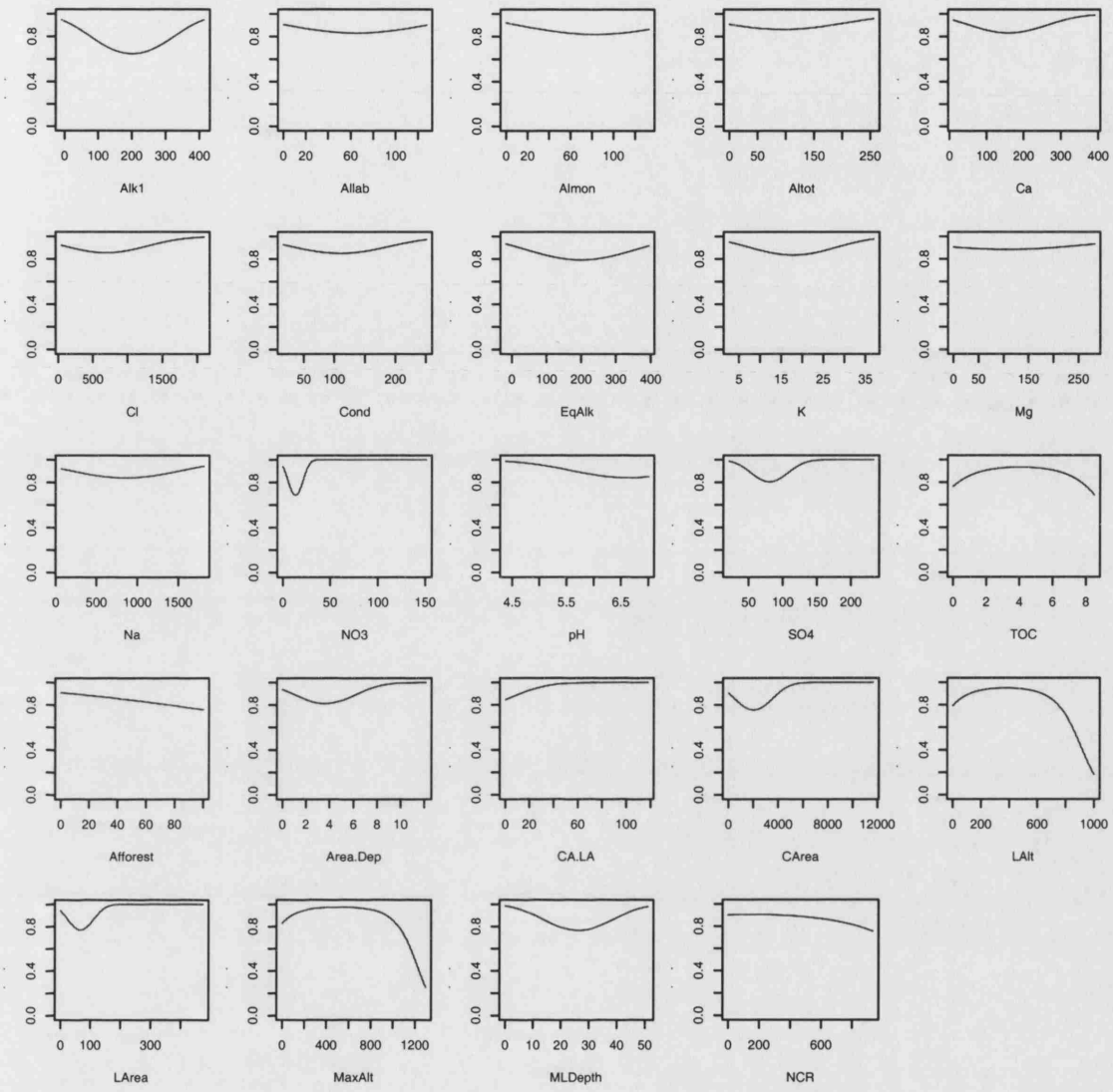


Figure 95: Generalised linear models of the response of Alonella nana to physicochemical parameters. The y-axis is the modelled probability of occurrence.

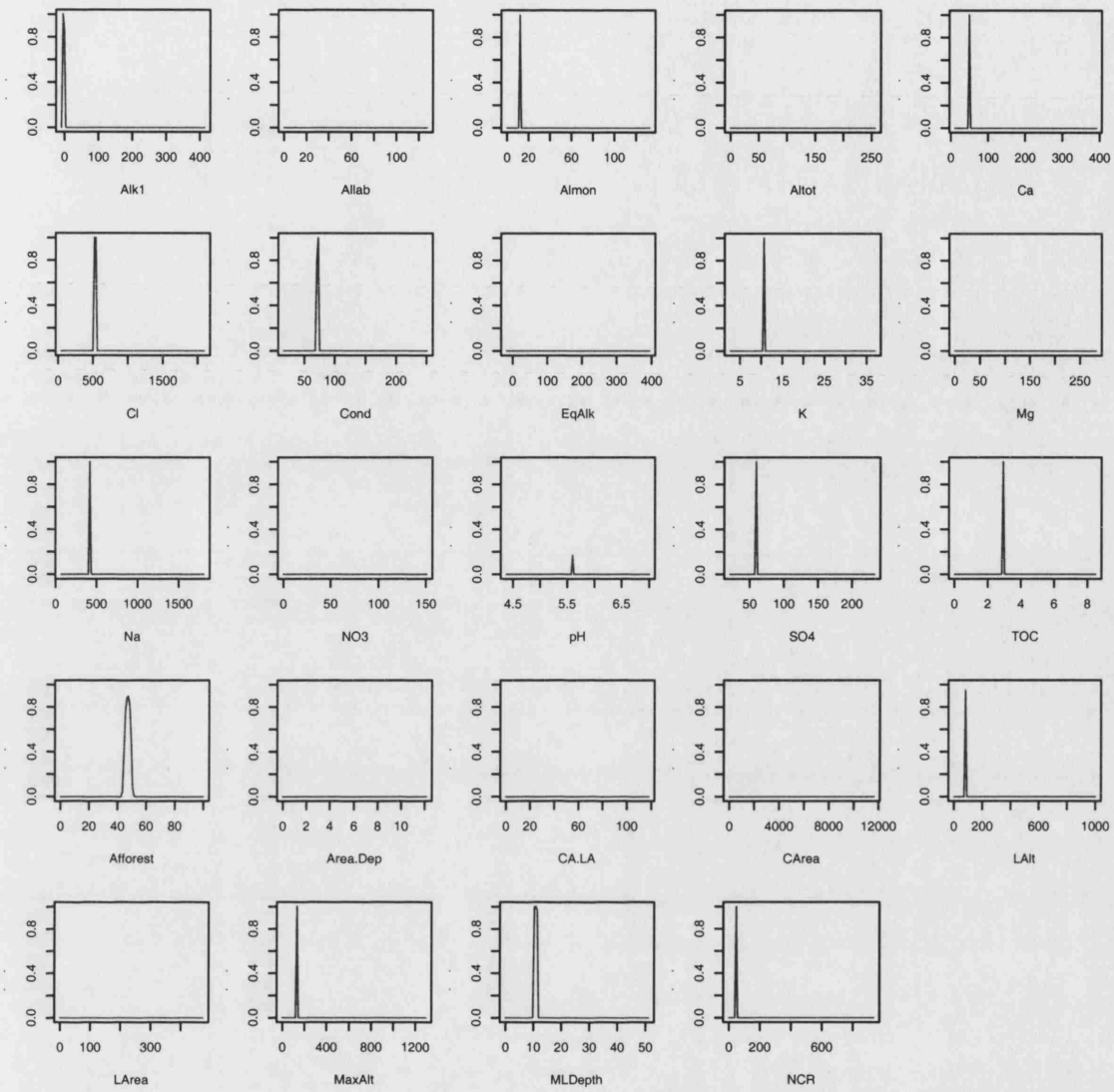


Figure 96: Generalised linear models of the response of Alonella rostrata to physicochemical parameters. The y-axis is the modelled probability of occurrence.

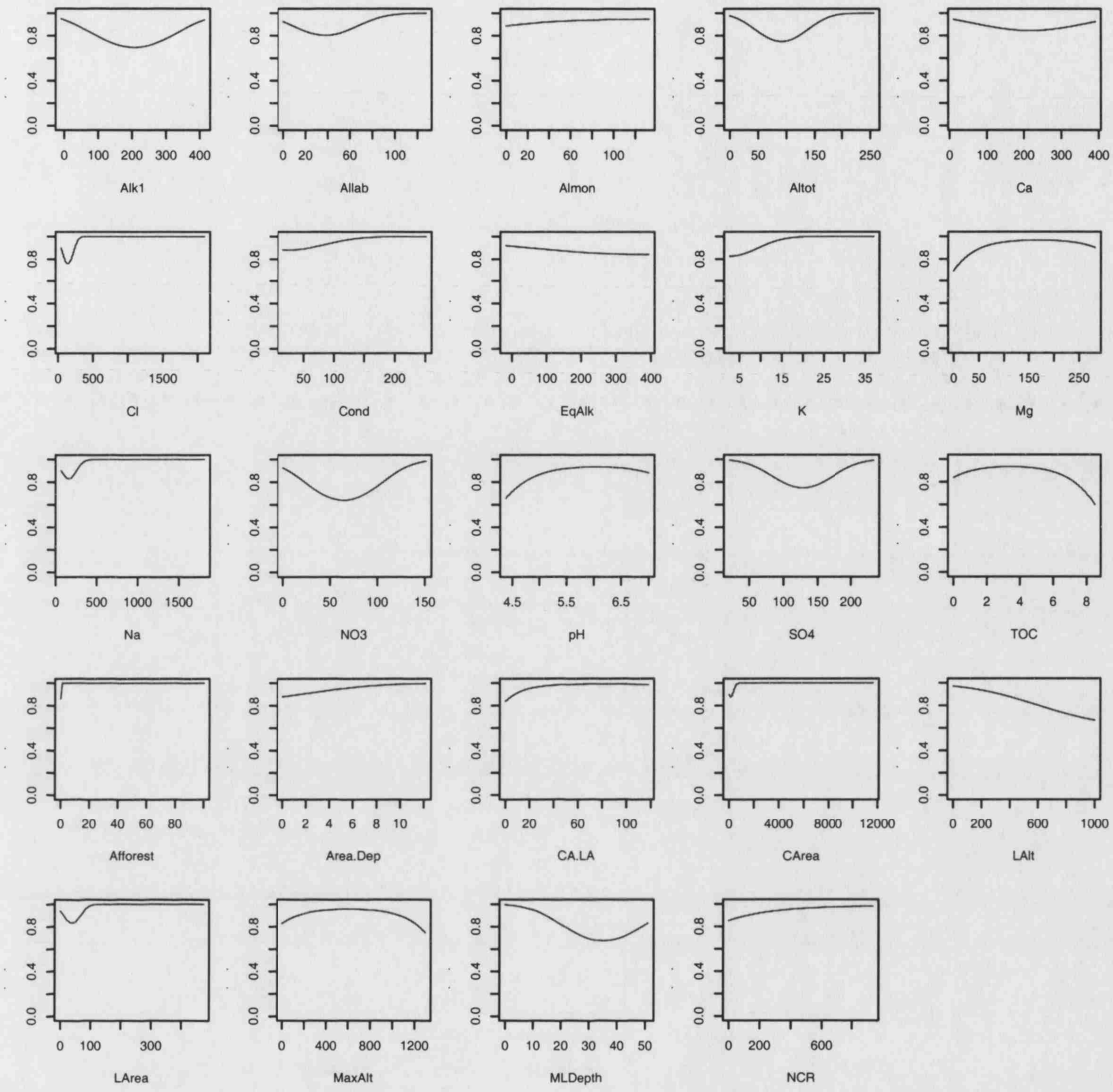


Figure 97: Generalised linear models of the response of Alona affinis to physicochemical parameters. The y-axis is the modelled probability of occurrence.

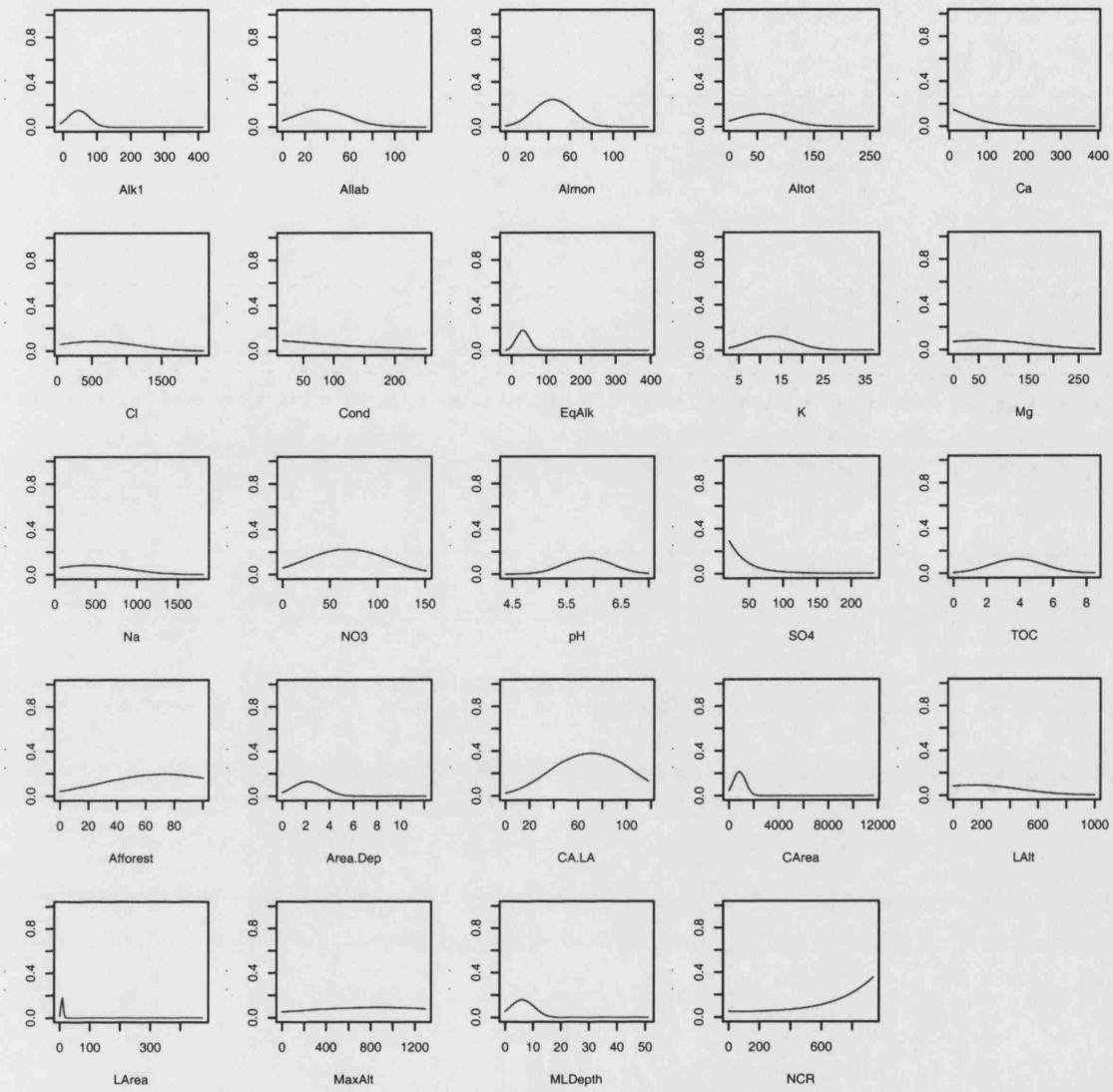


Figure 98: Generalised linear models of the response of Alona costata to physicochemical parameters. The y-axis is the modelled probability of occurrence.

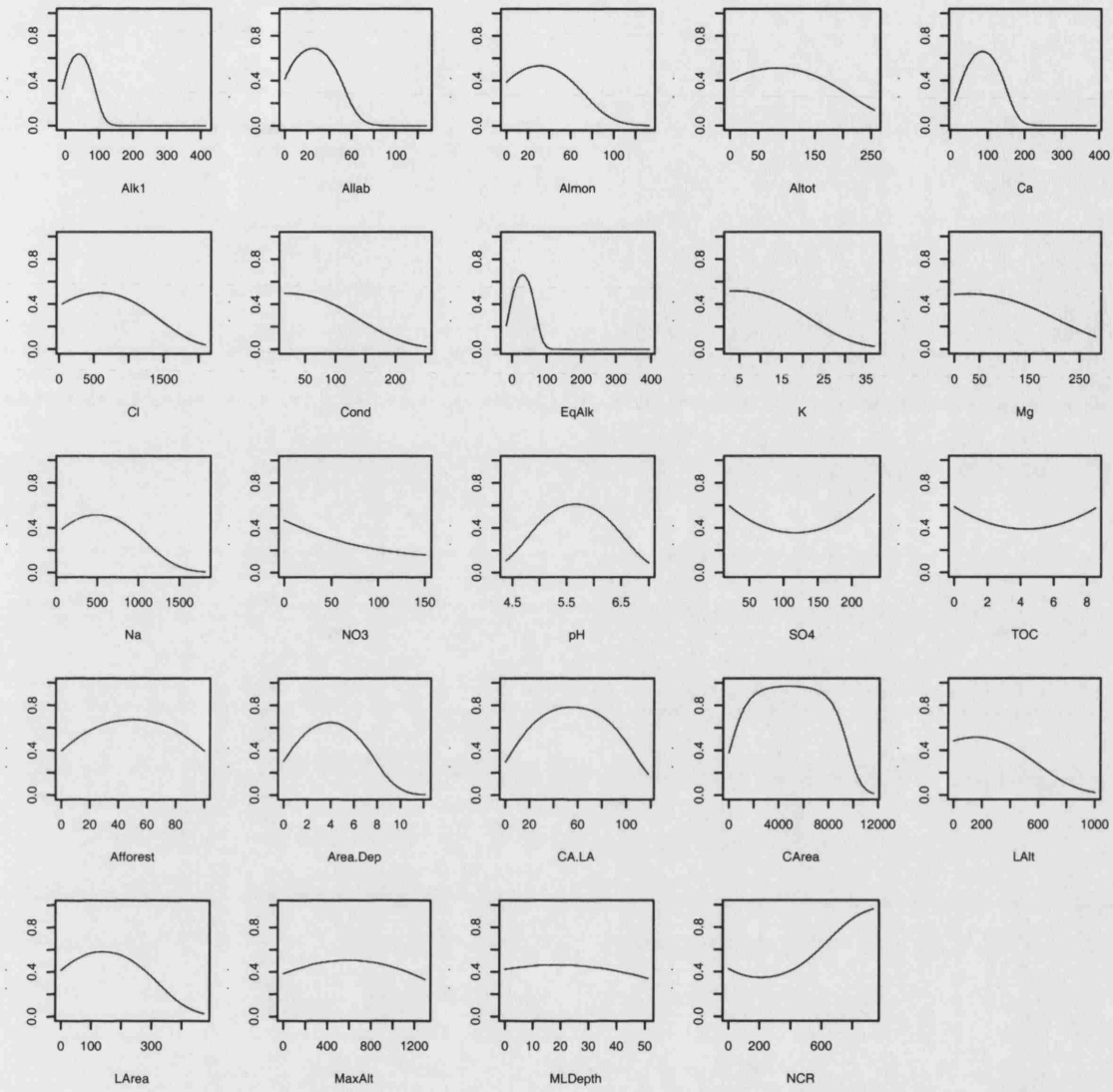


Figure 99: Generalised linear models of the response of Alona guttata var. tuberculata to physicochemical parameters. The y-axis is the modelled probability of occurrence.

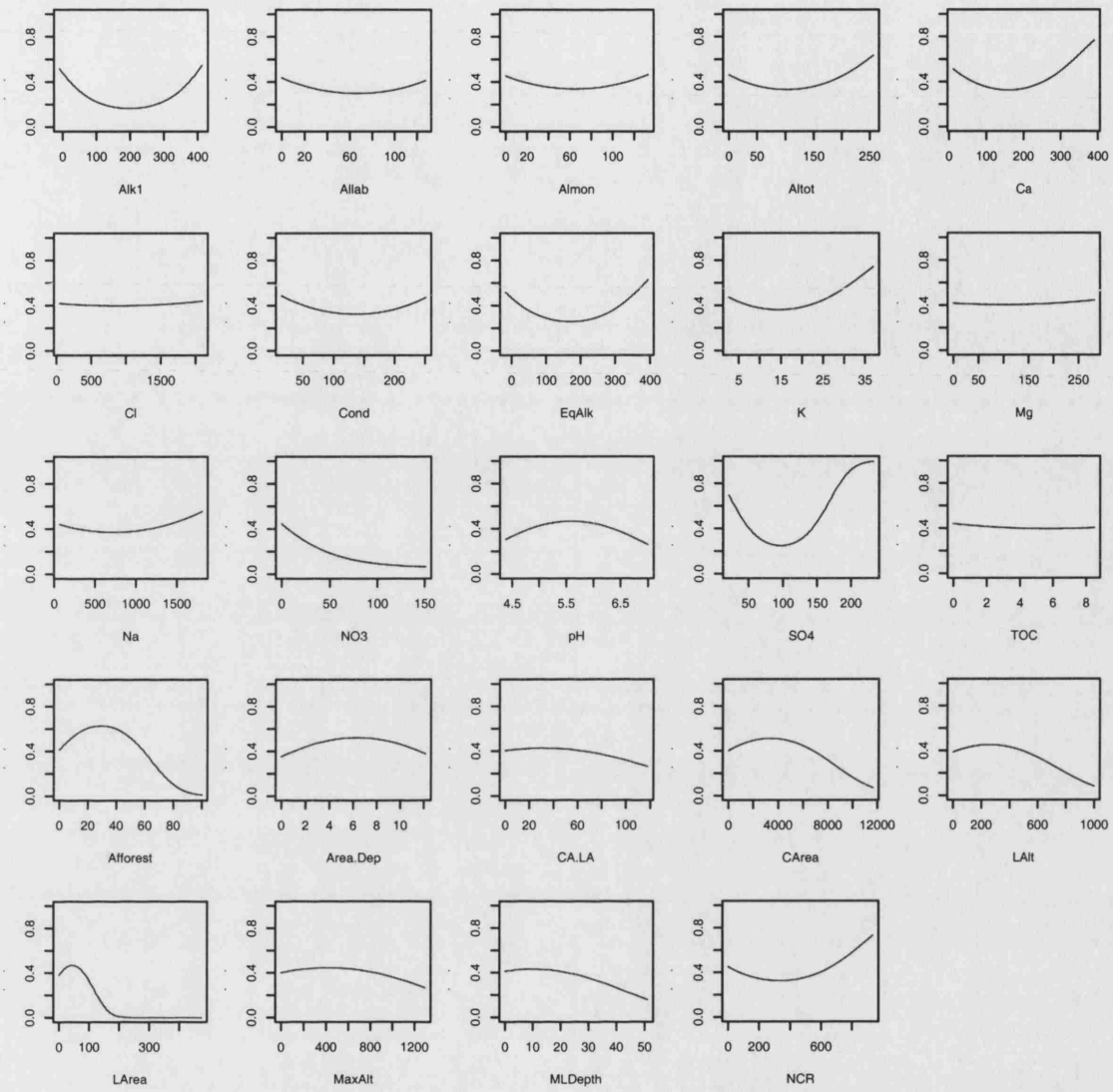


Figure 100: Generalised linear models of the response of Alona guttata to physicochemical parameters. The y-axis is the modelled probability of occurrence.

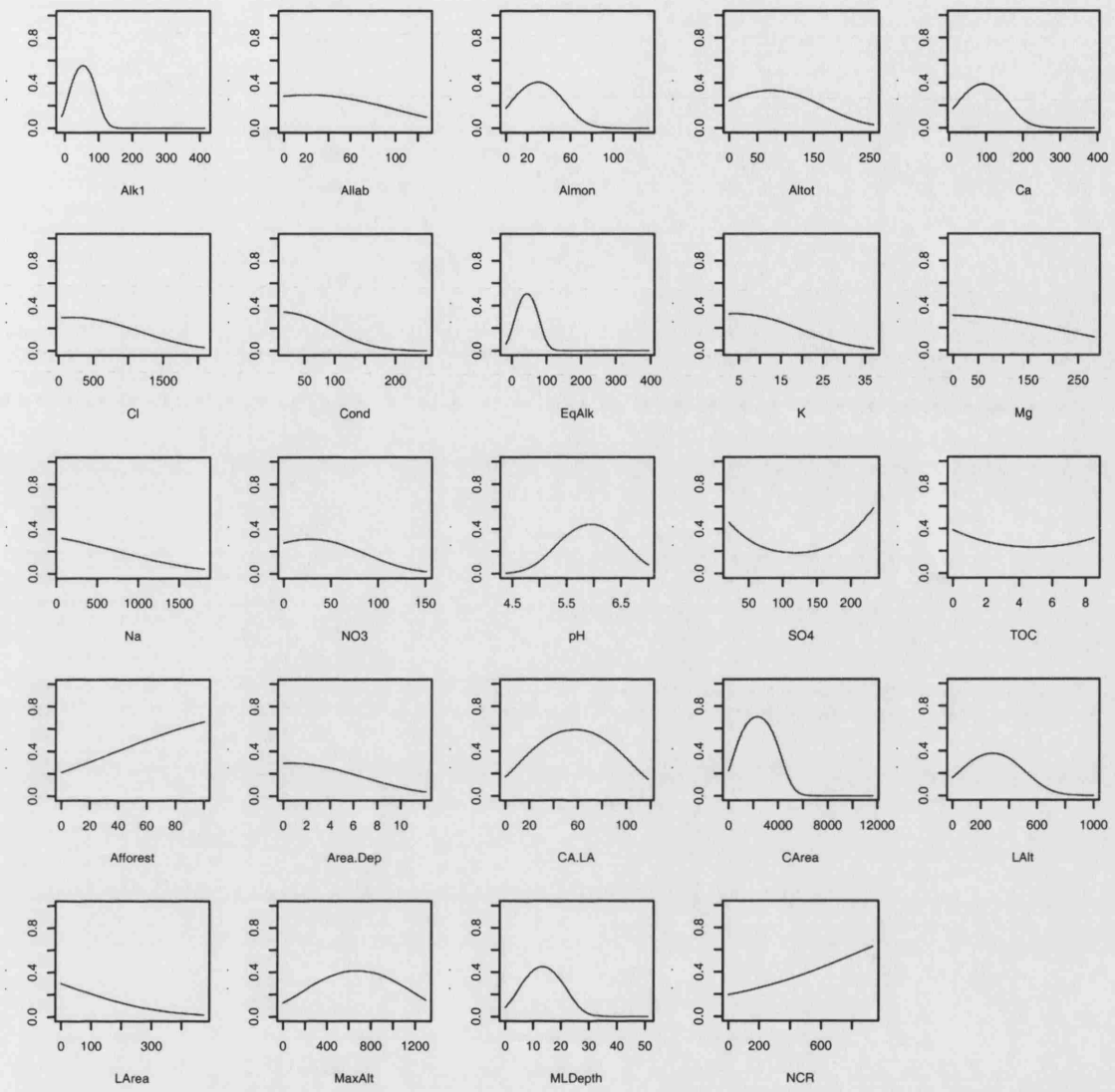


Figure 101: Generalised linear models of the response of Alona intemedia to physicochemical parameters. The y-axis is the modelled probability of occurrence.

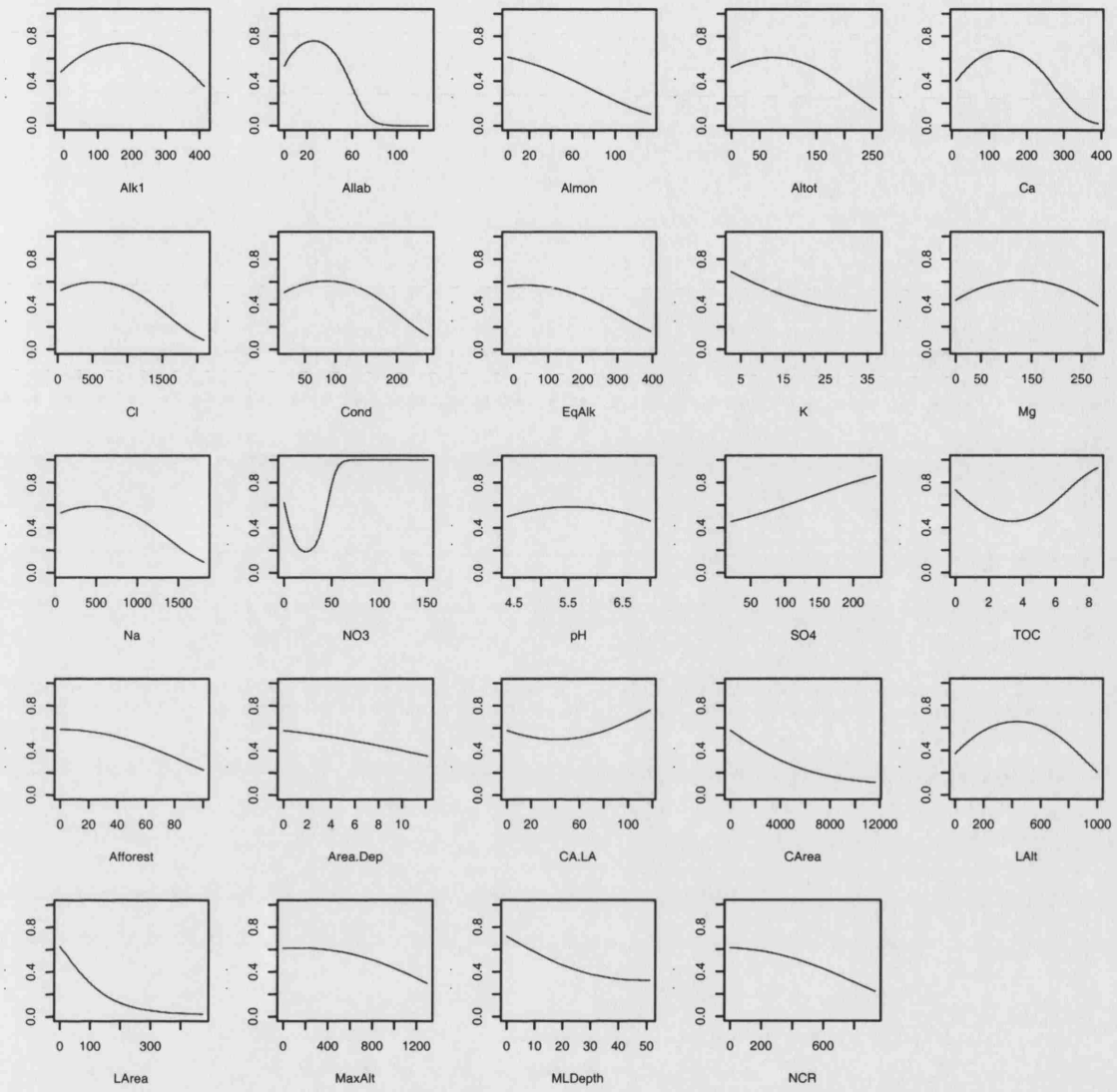


Figure 102: Generalised linear models of the response of Alona quadrangularis to physicochemical parameters. The y-axis is the modelled probability of occurrence.

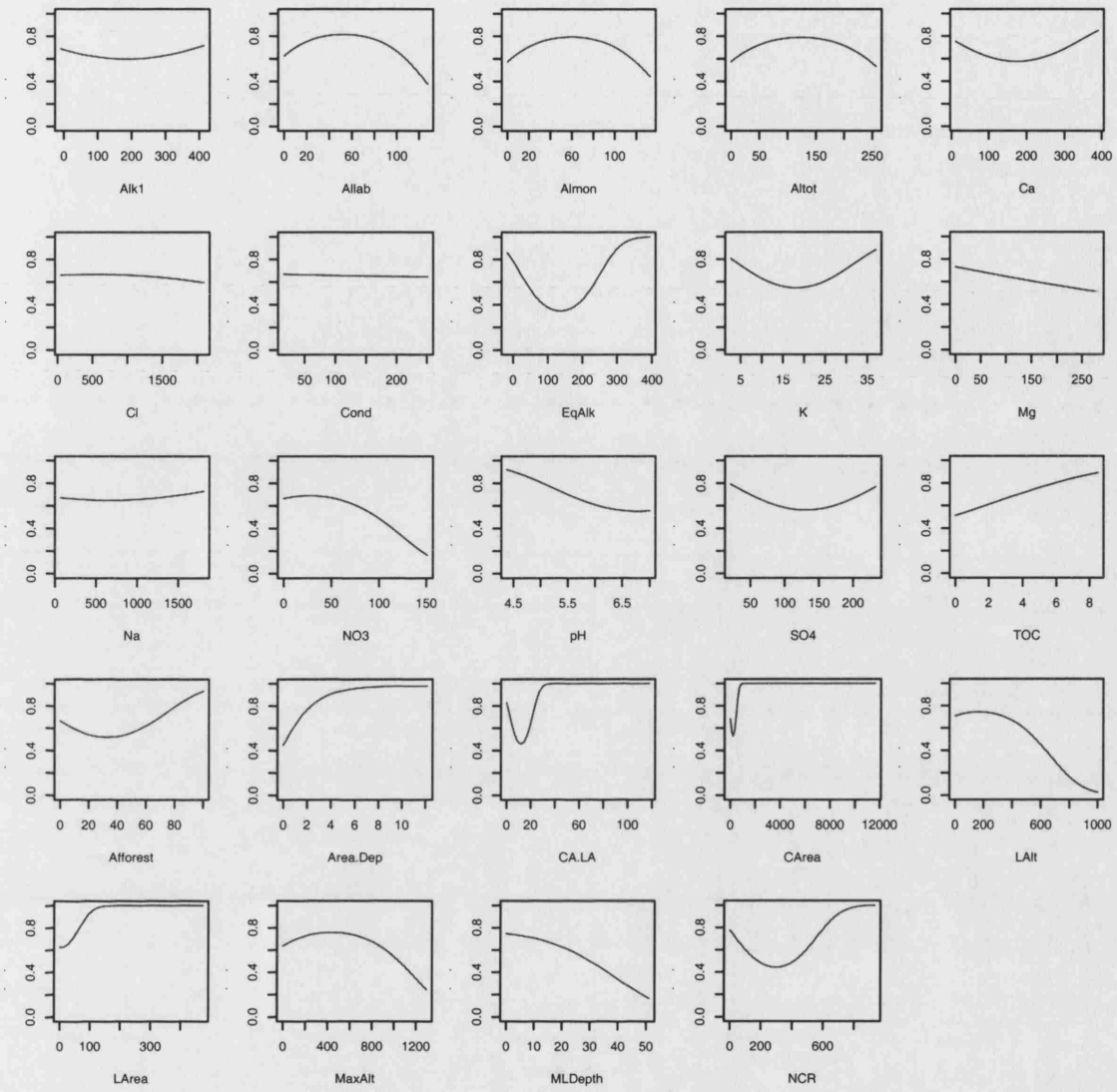


Figure 103: Generalised linear models of the response of Alona rectangula to physicochemical parameters. The y-axis is the modelled probability of occurrence.

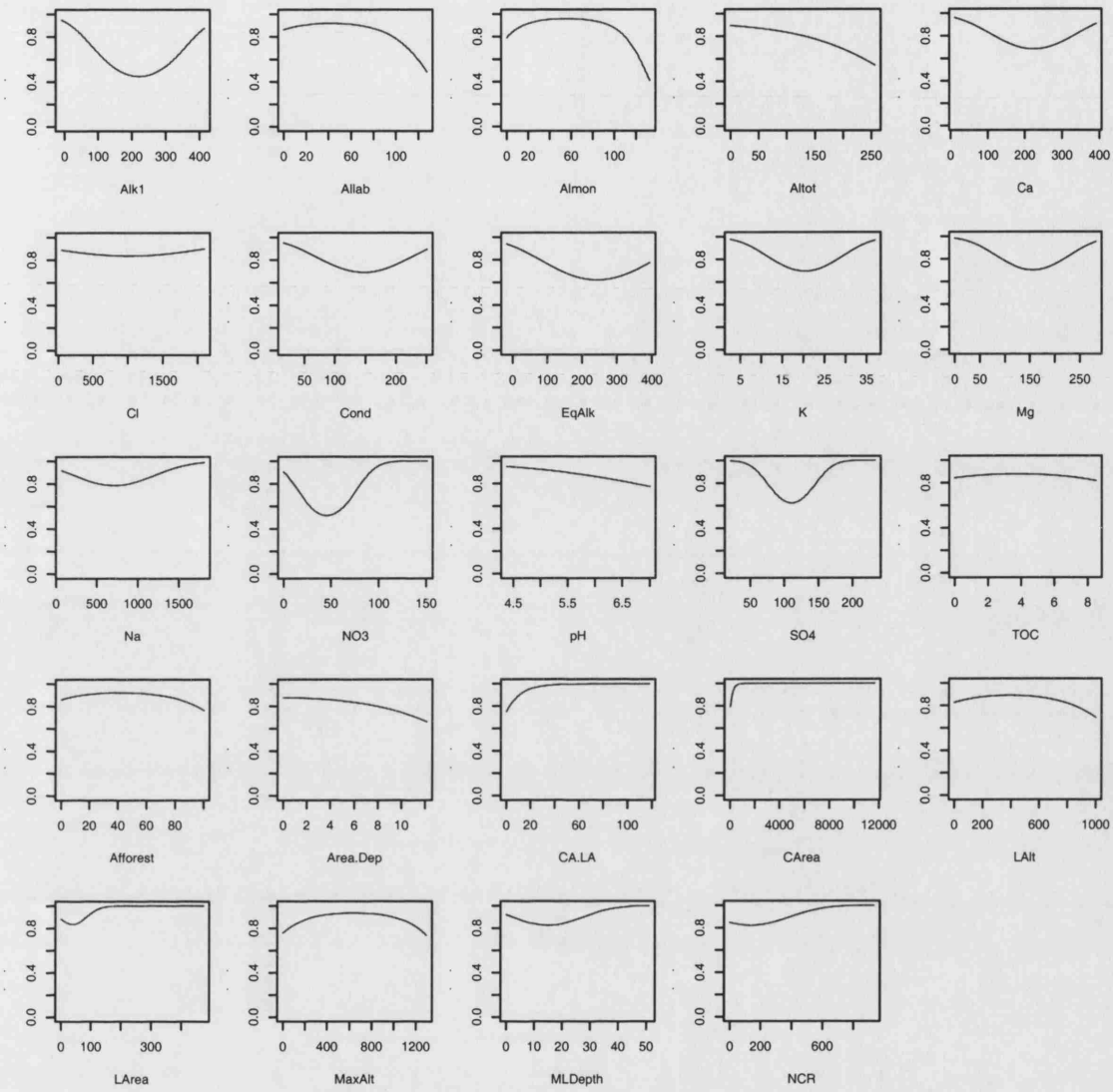


Figure 104: Generalised linear models of the response of Alona rustica to physicochemical parameters. The y-axis is the modelled probability of occurrence.

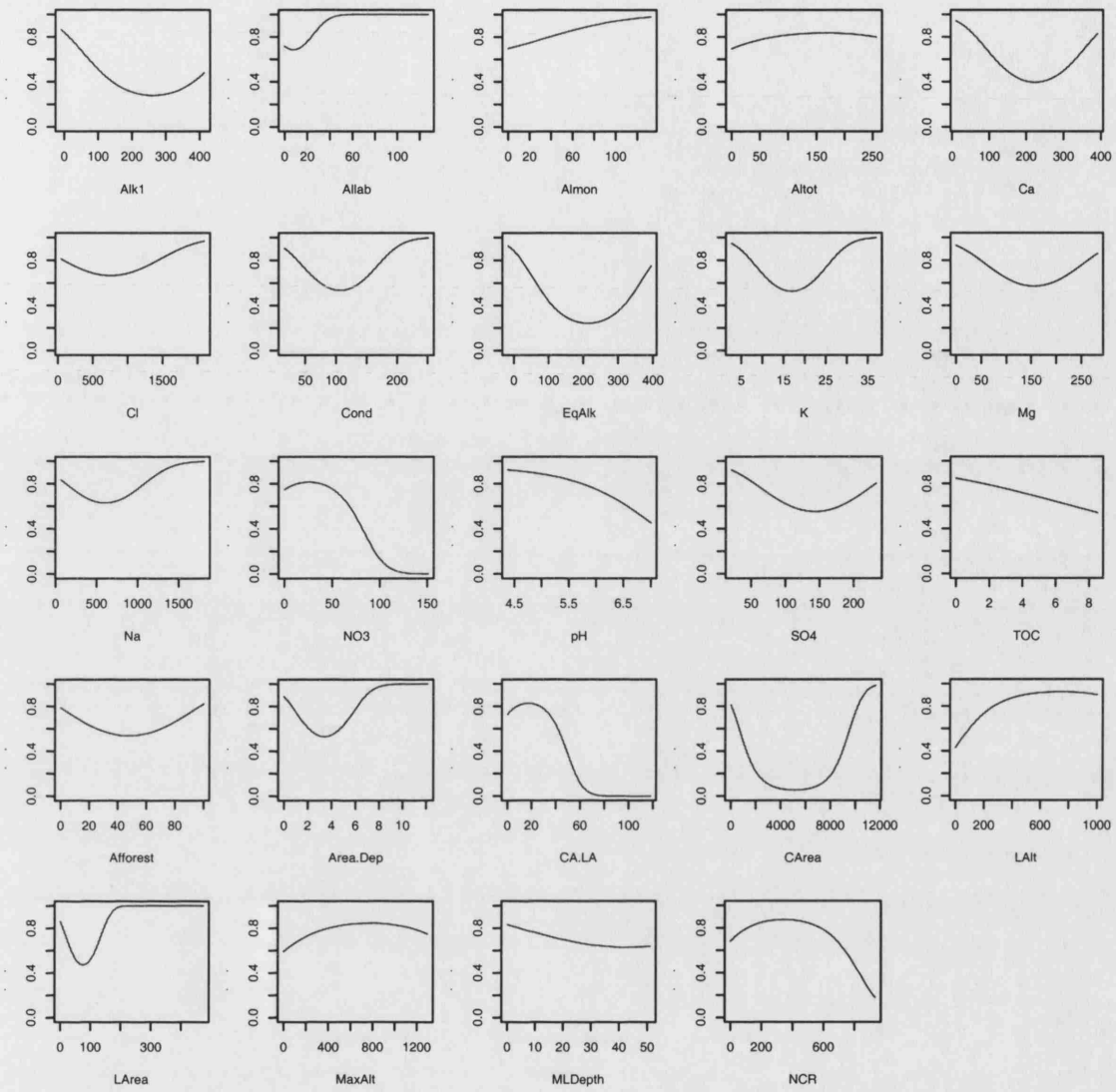


Figure 105: Generalised linear models of the response of Alonopsis elongatus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

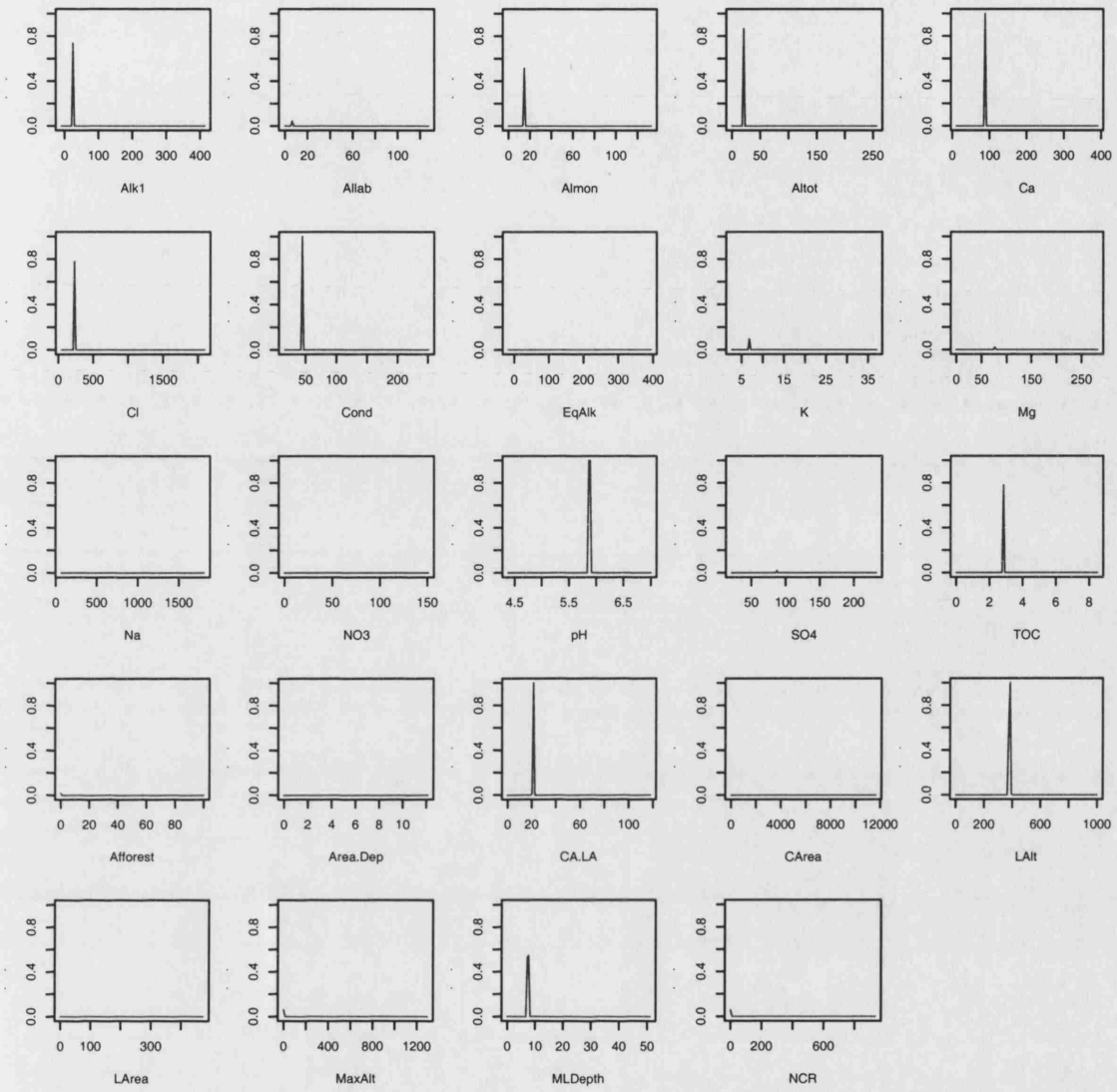


Figure 106: Generalised linear models of the response of Anchistropus emarginatus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

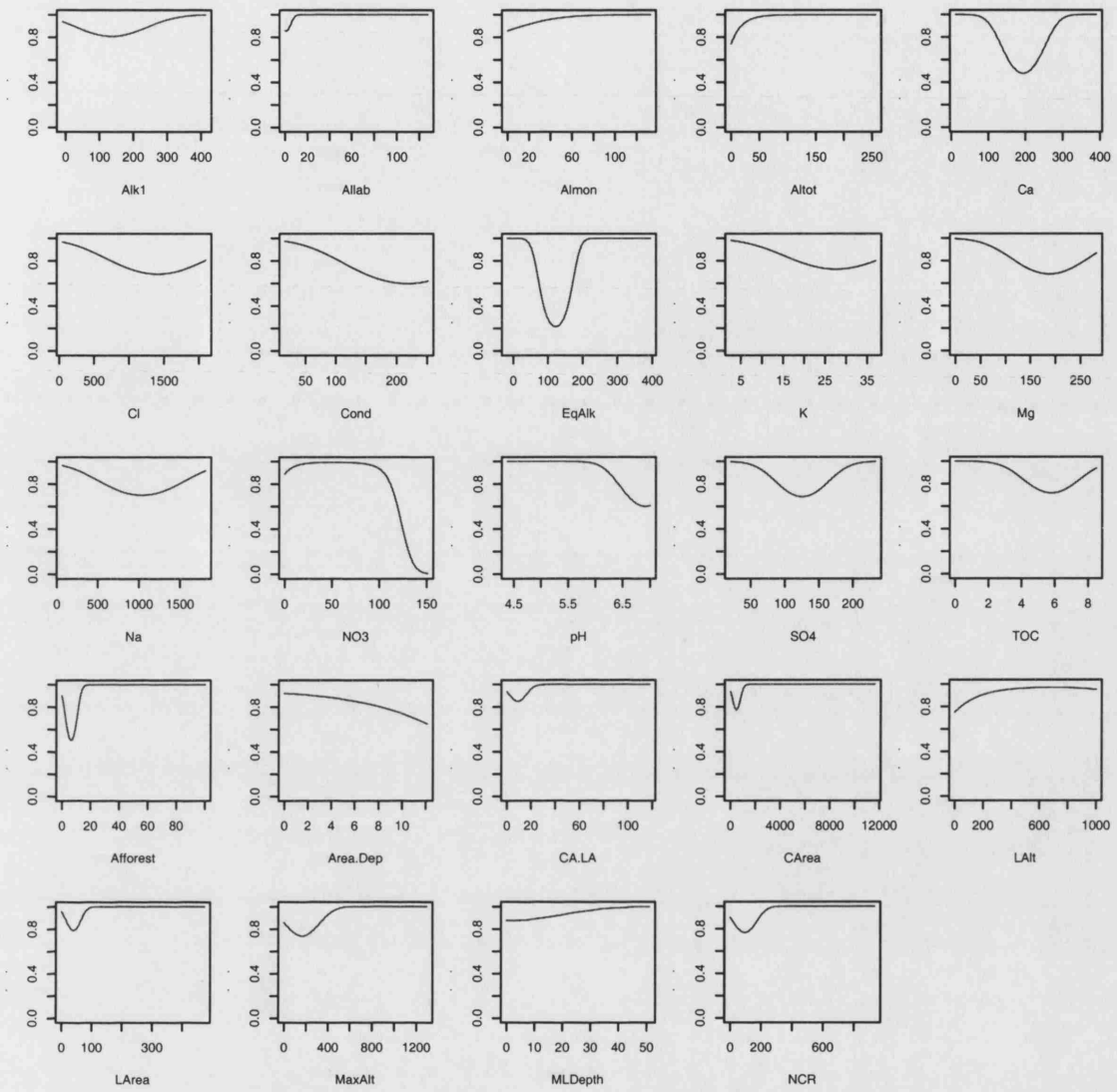


Figure 107: Generalised linear models of the response of Bosmina coregoni to physicochemical parameters. The y-axis is the modelled probability of occurrence.

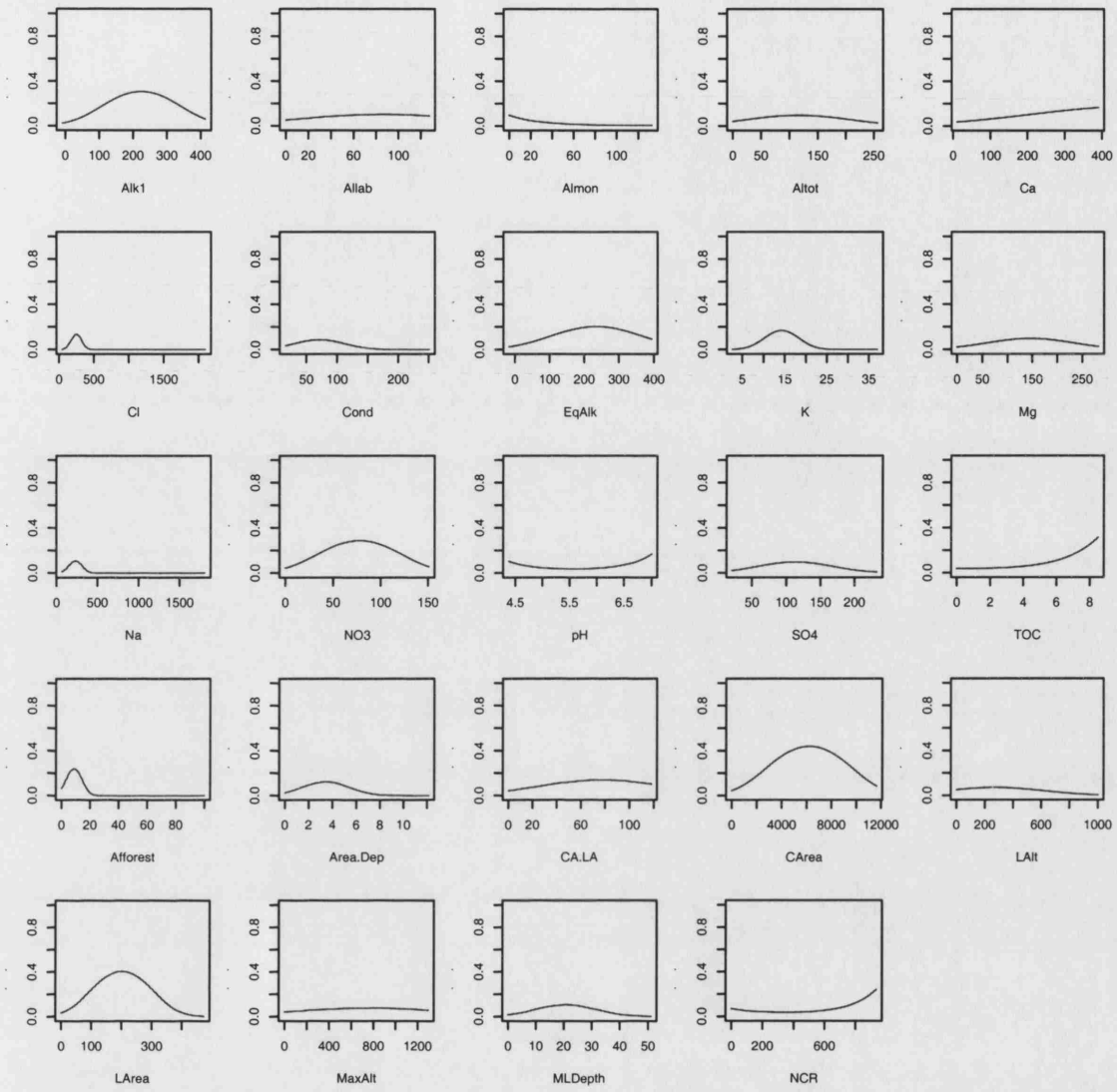


Figure 108: Generalised linear models of the response of Bosmina longirostris to physicochemical parameters. The y-axis is the modelled probability of occurrence.

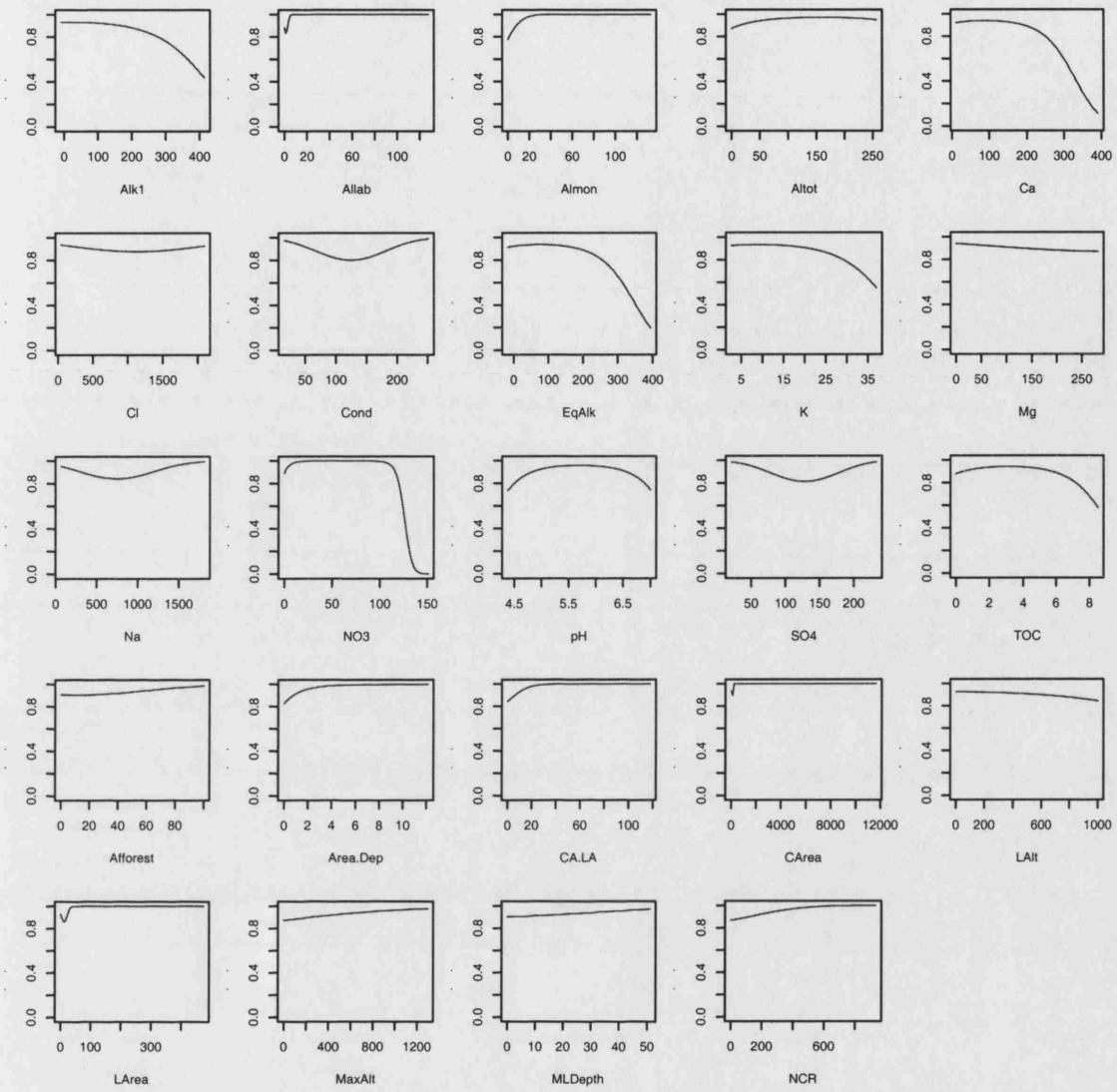


Figure 109: Generalised linear models of the response of Bosmina longispina to physicochemical parameters. The y-axis is the modelled probability of occurrence.

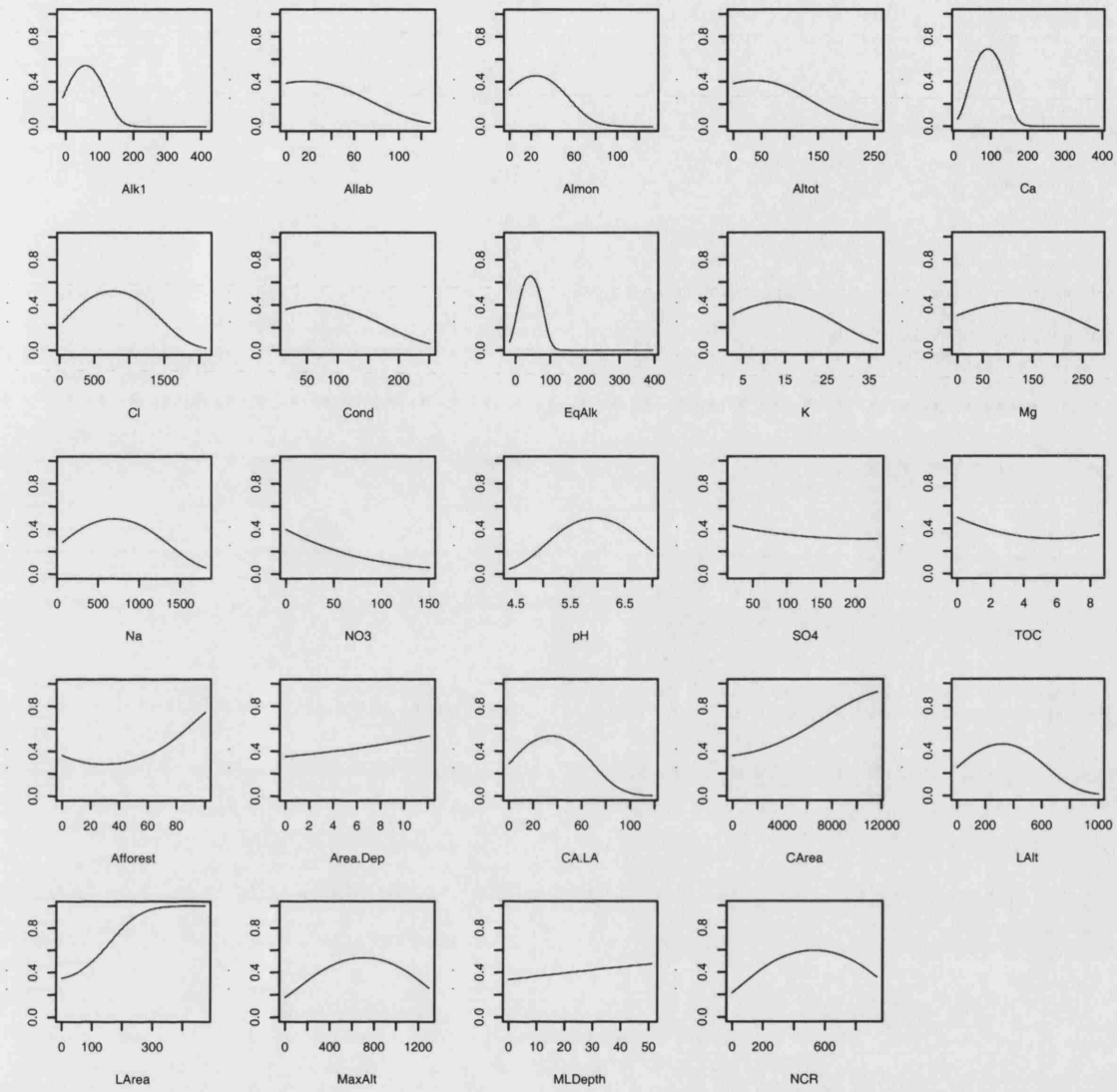


Figure 110: Generalised linear models of the response of Bosmina spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

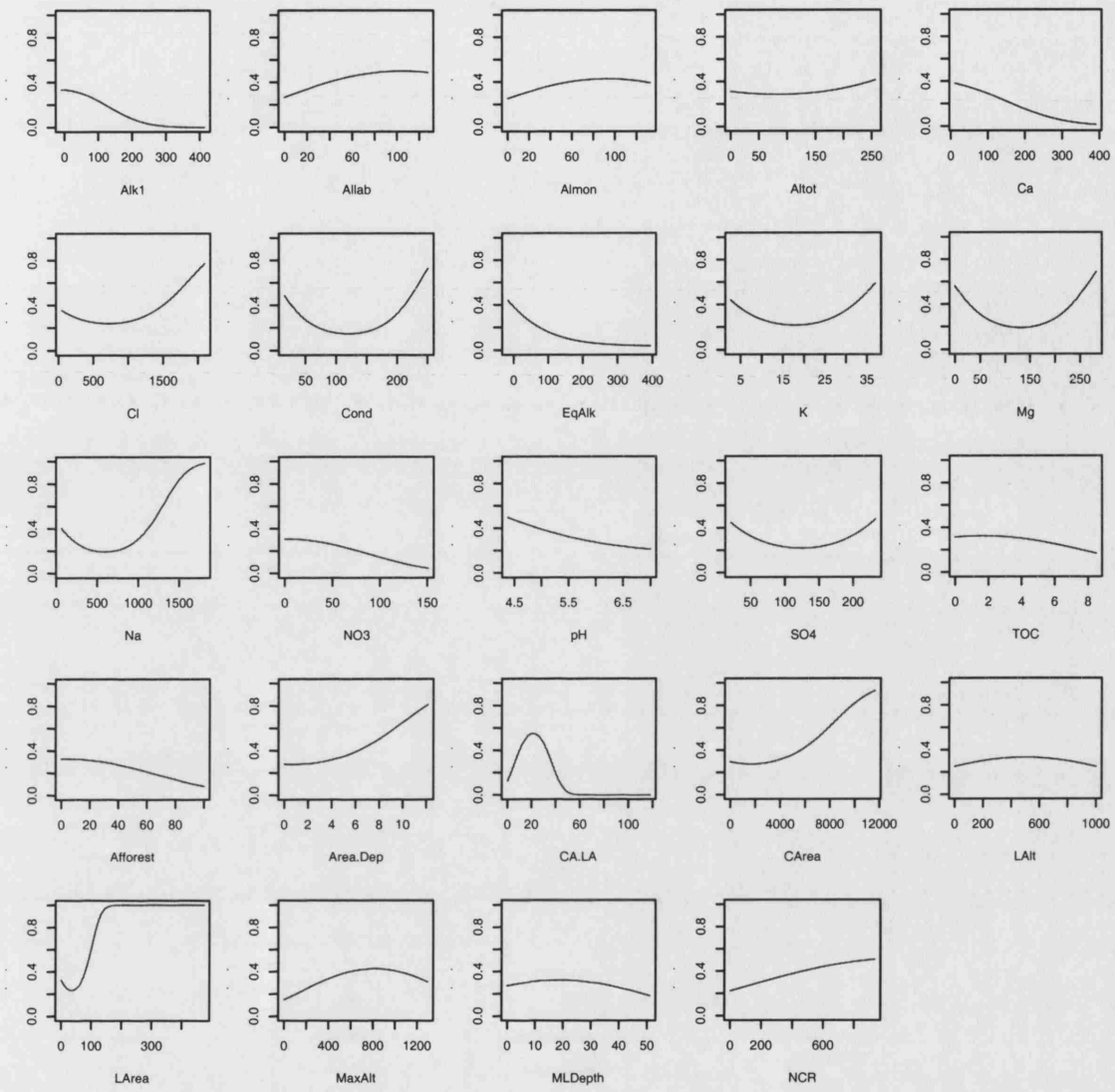


Figure 111: Generalised linear models of the response of Camptocercus rectirotris to physicochemical parameters. The y-axis is the modelled probability of occurrence.

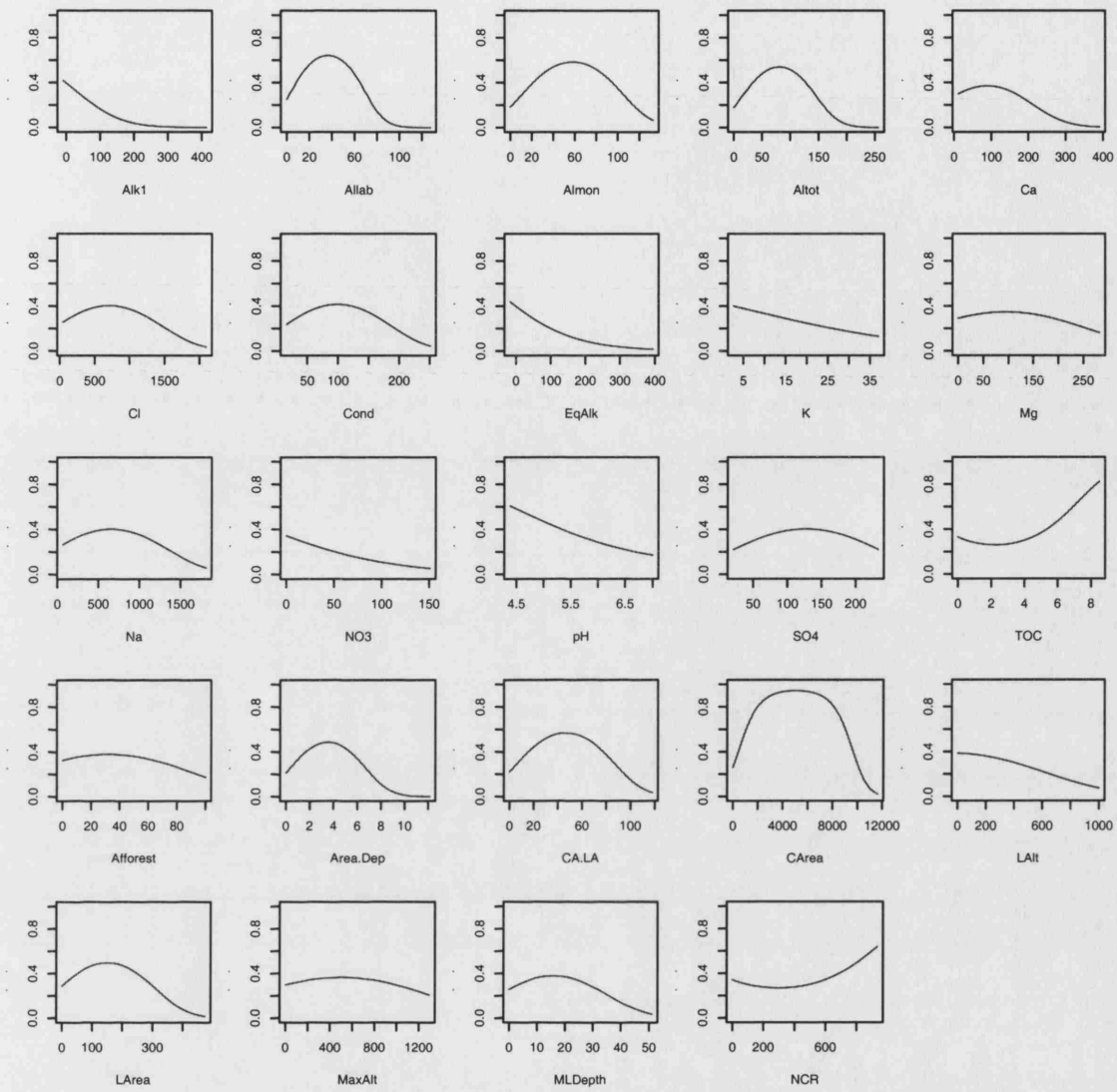


Figure 112: Generalised linear models of the response of Ceriodaphnia spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

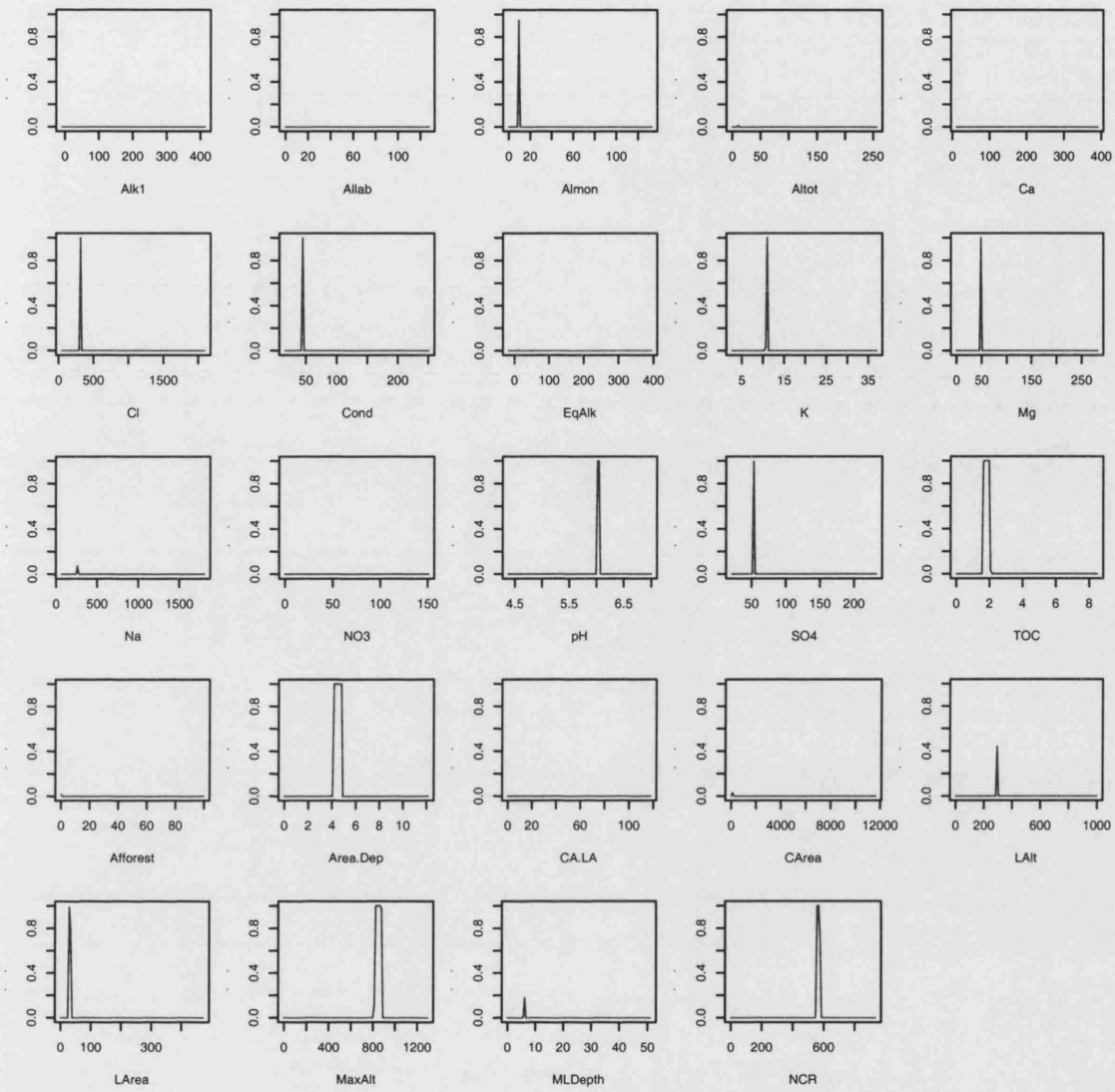


Figure 113: Generalised linear models of the response of Chydorus gibbus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

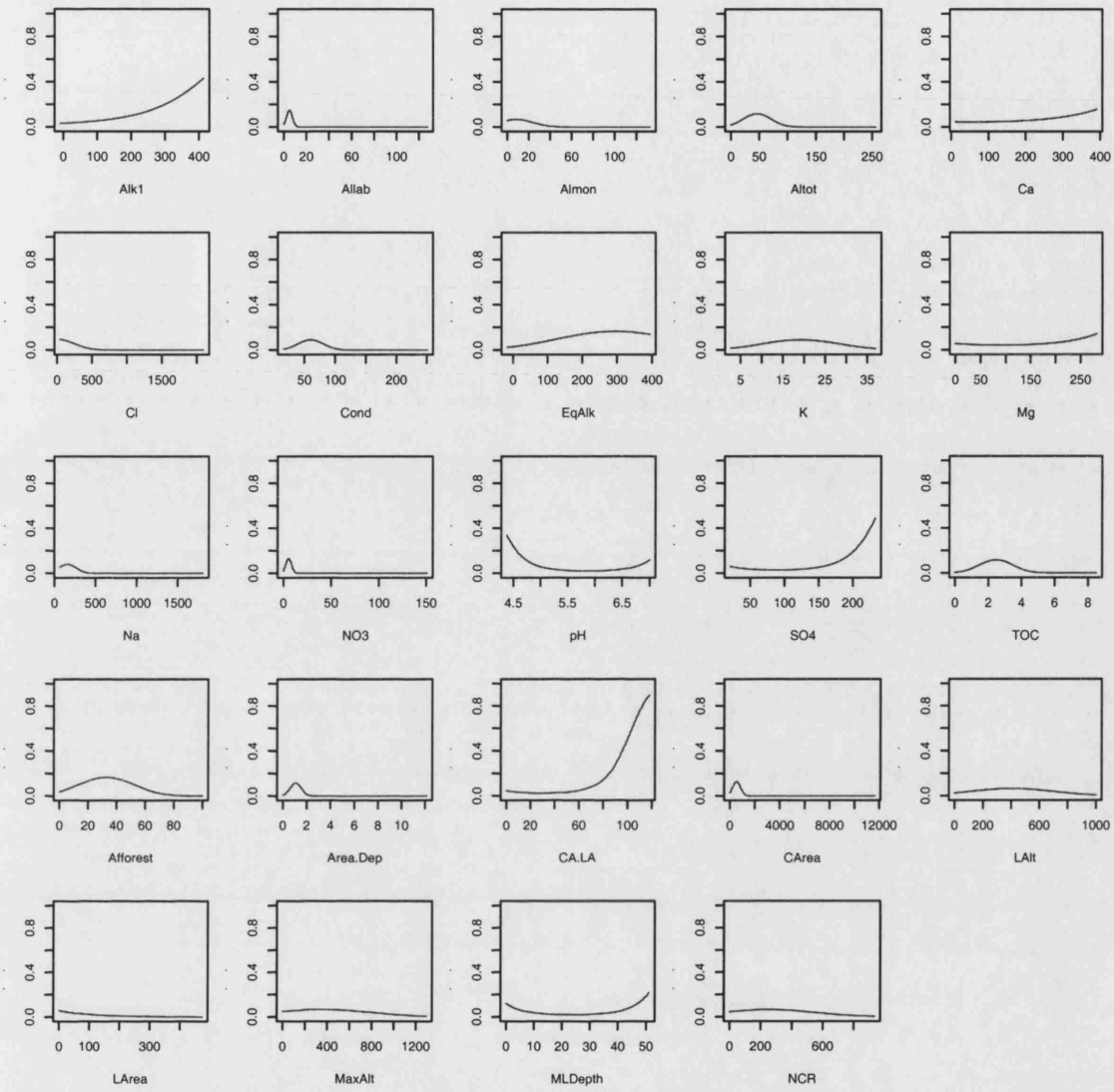


Figure 114: Generalised linear models of the response of Chydorus ovalis to physicochemical parameters. The y-axis is the modelled probability of occurrence.

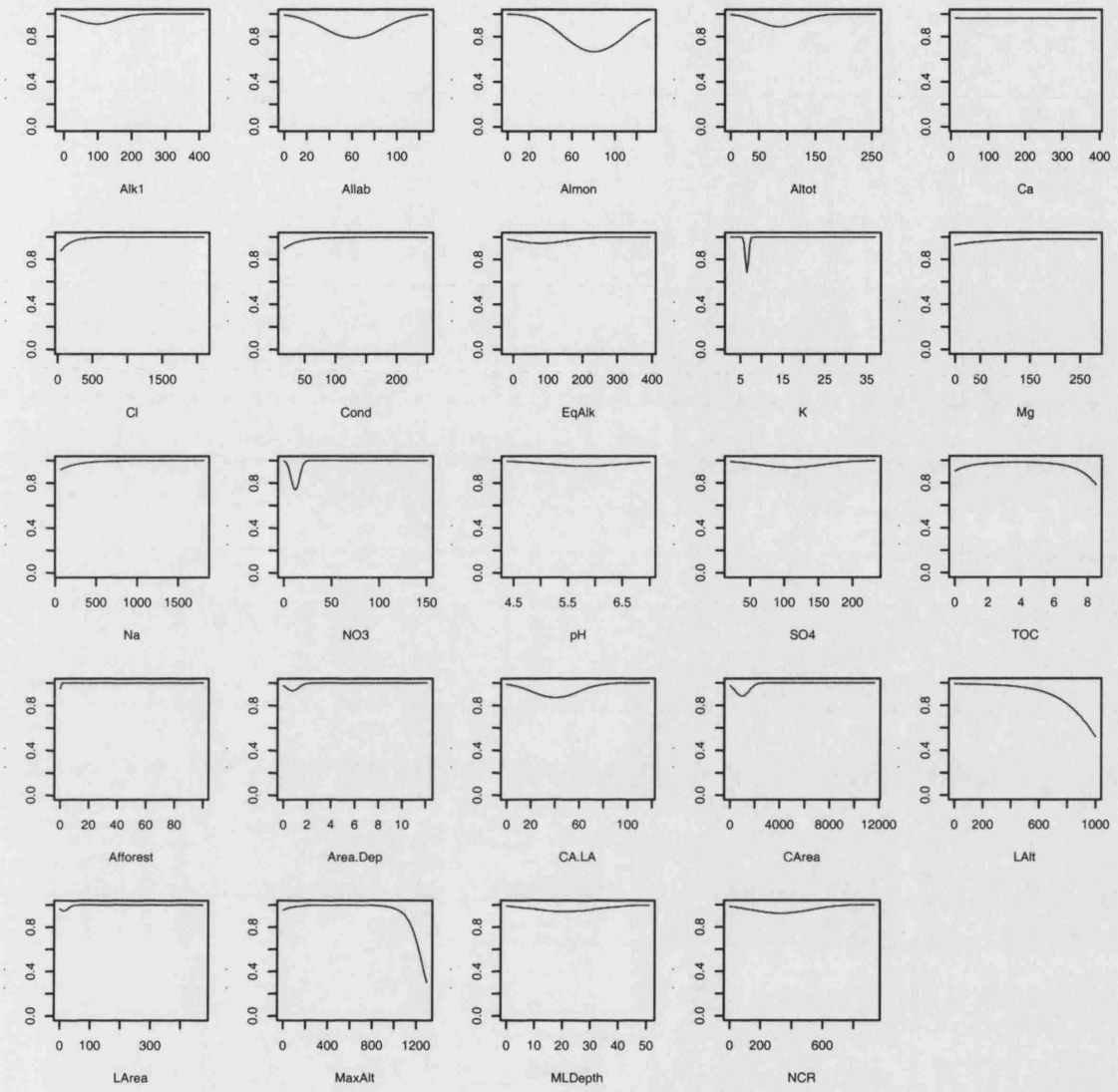


Figure 115: Generalised linear models of the response of Chydorus piger to physicochemical parameters. The y-axis is the modelled probability of occurrence.

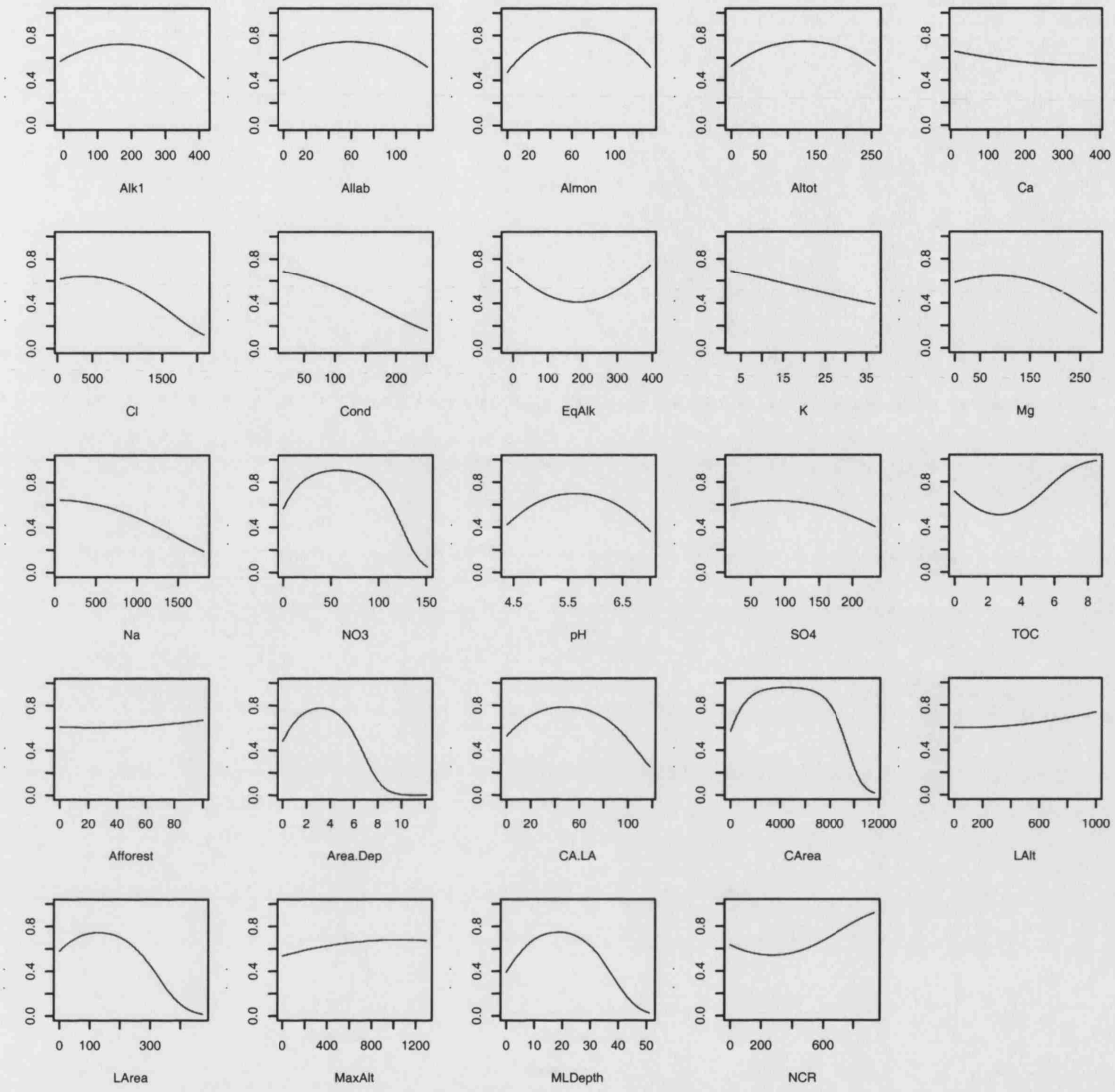


Figure 116: Generalised linear models of the response of Chydorus sphaericus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

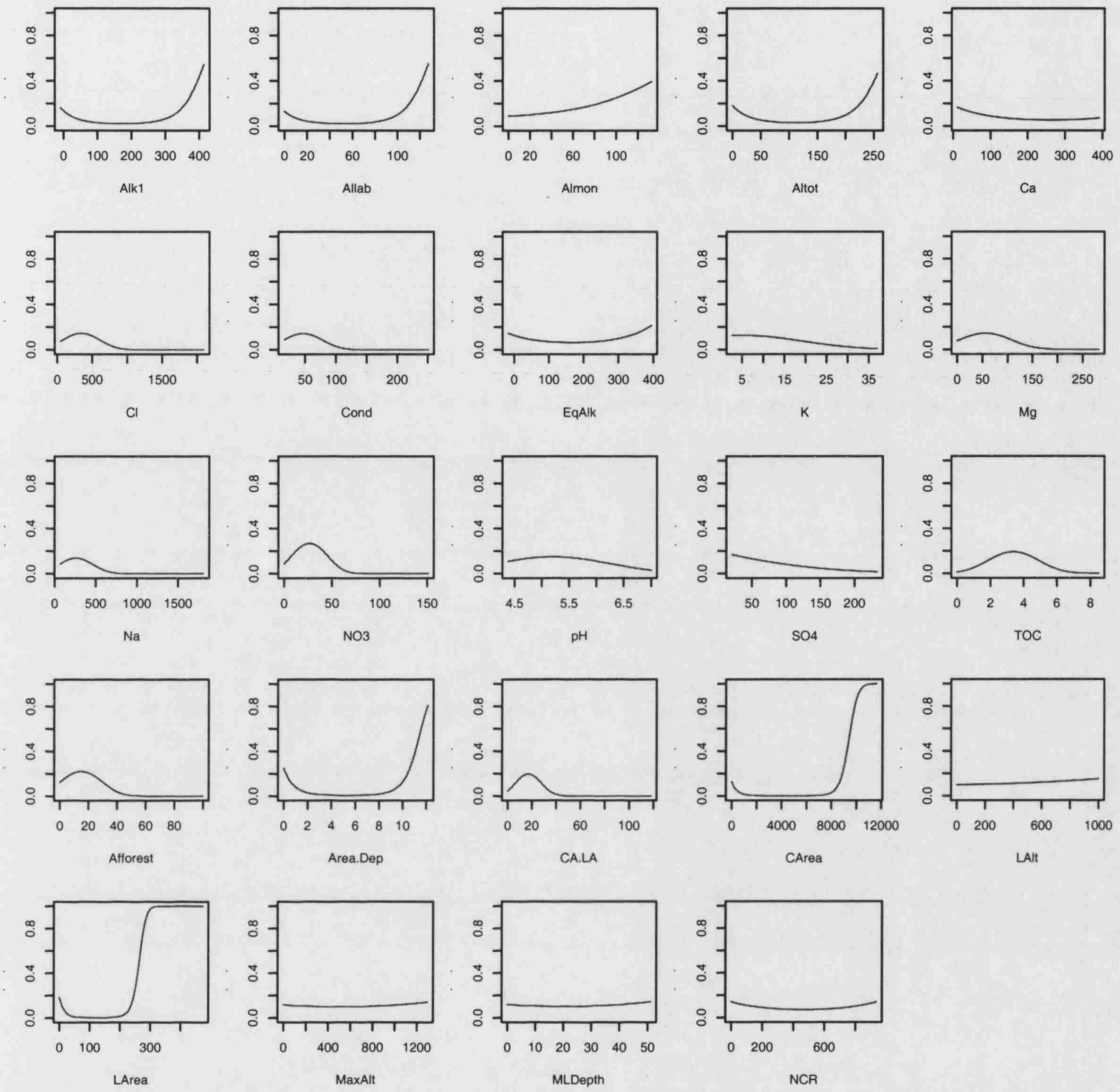


Figure 117: Generalised linear models of the response of Daphnia longispina grp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

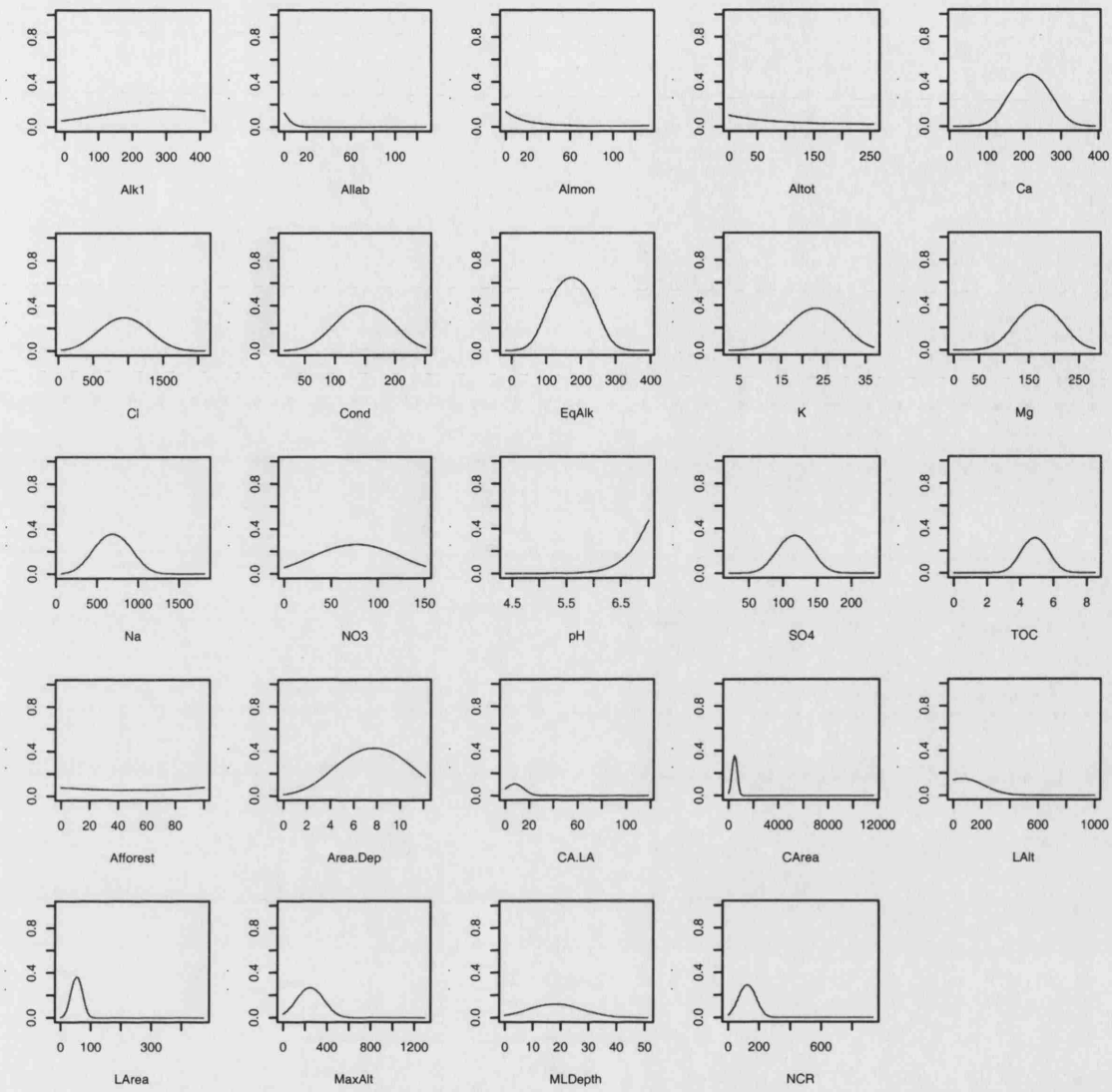


Figure 118: Generalised linear models of the response of Daphnia pulex grp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

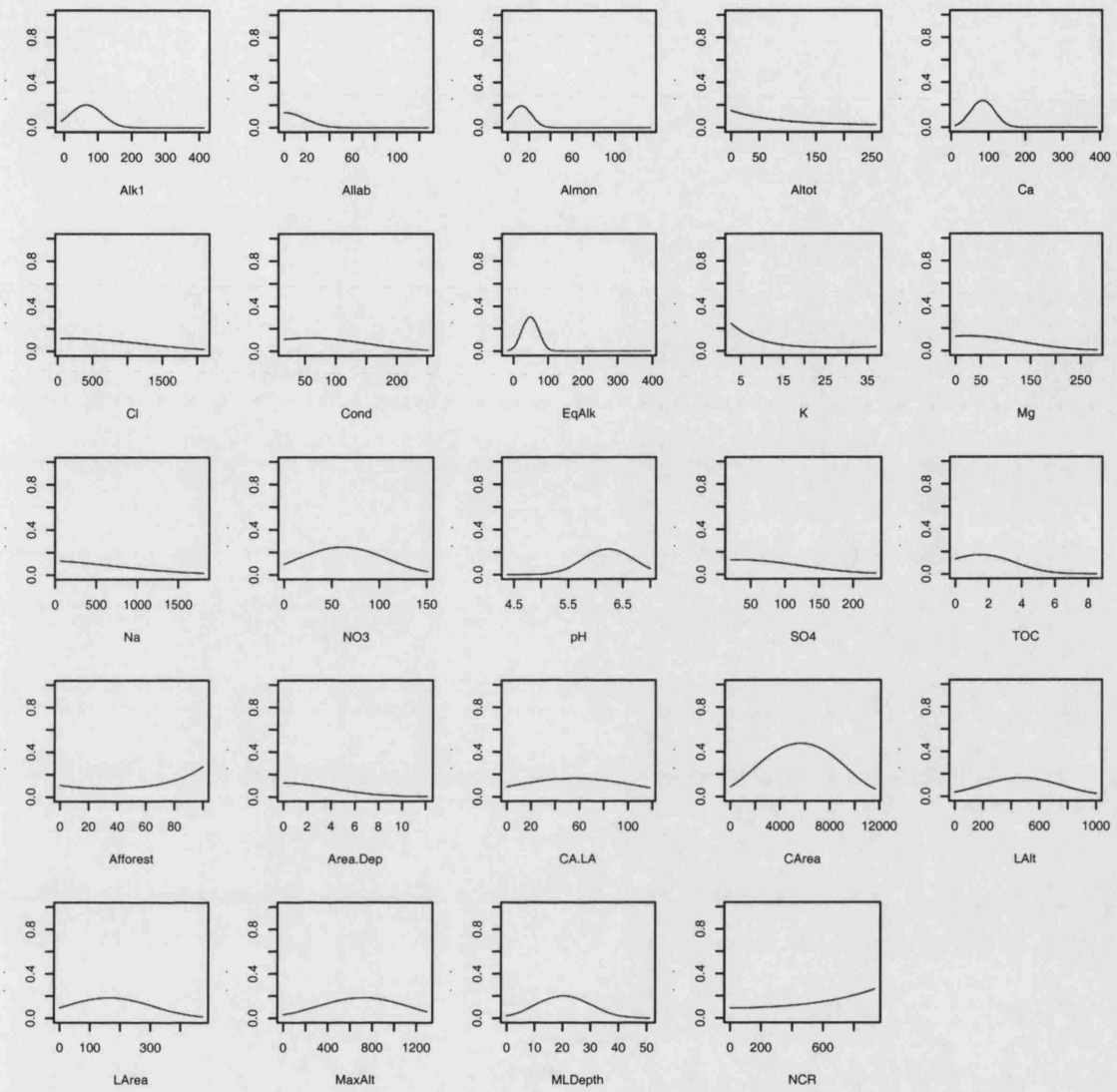


Figure 119: Generalised linear models of the response of Daphnia spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

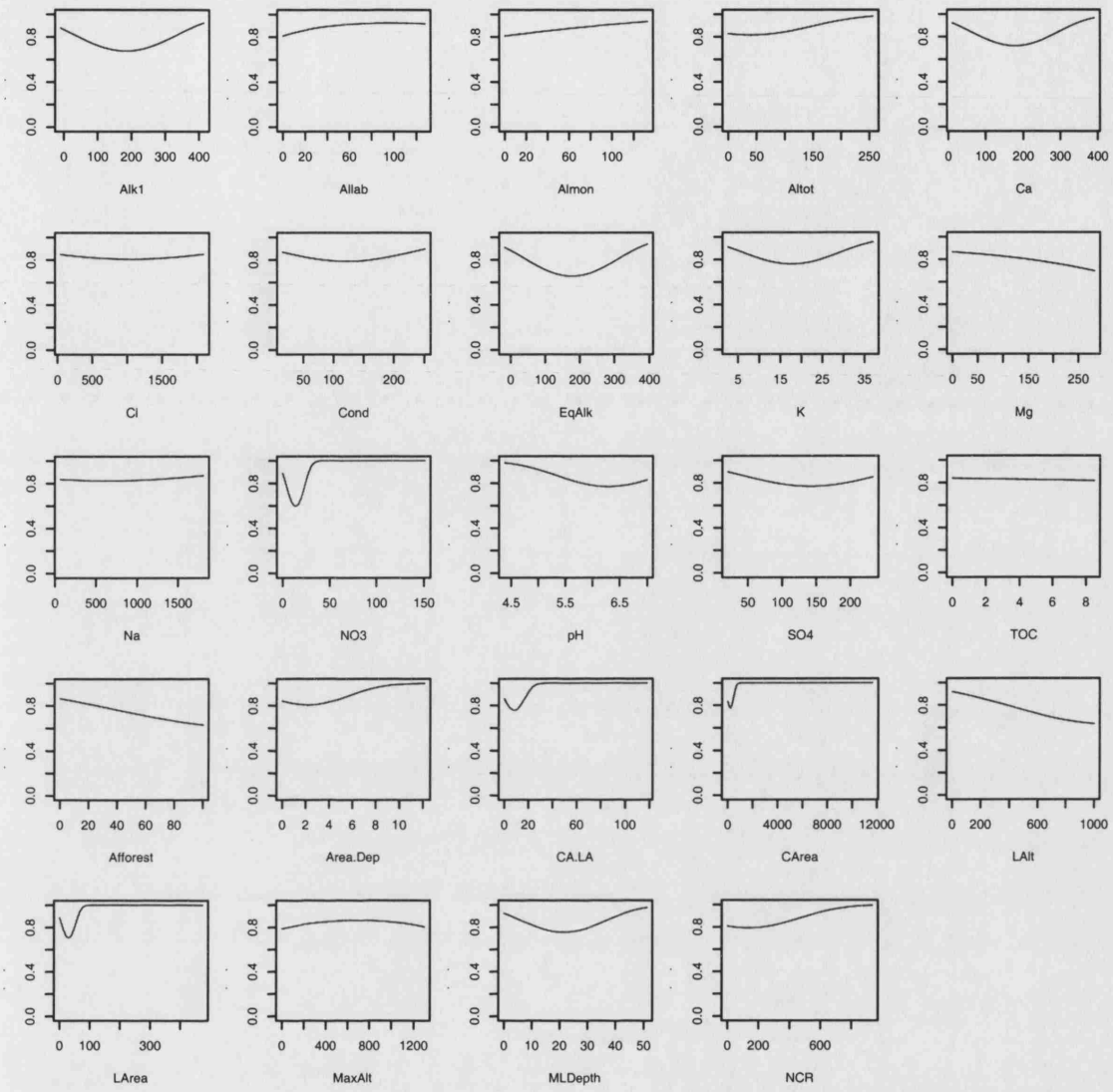


Figure 120: Generalised linear models of the response of Eurycercus lamellatus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

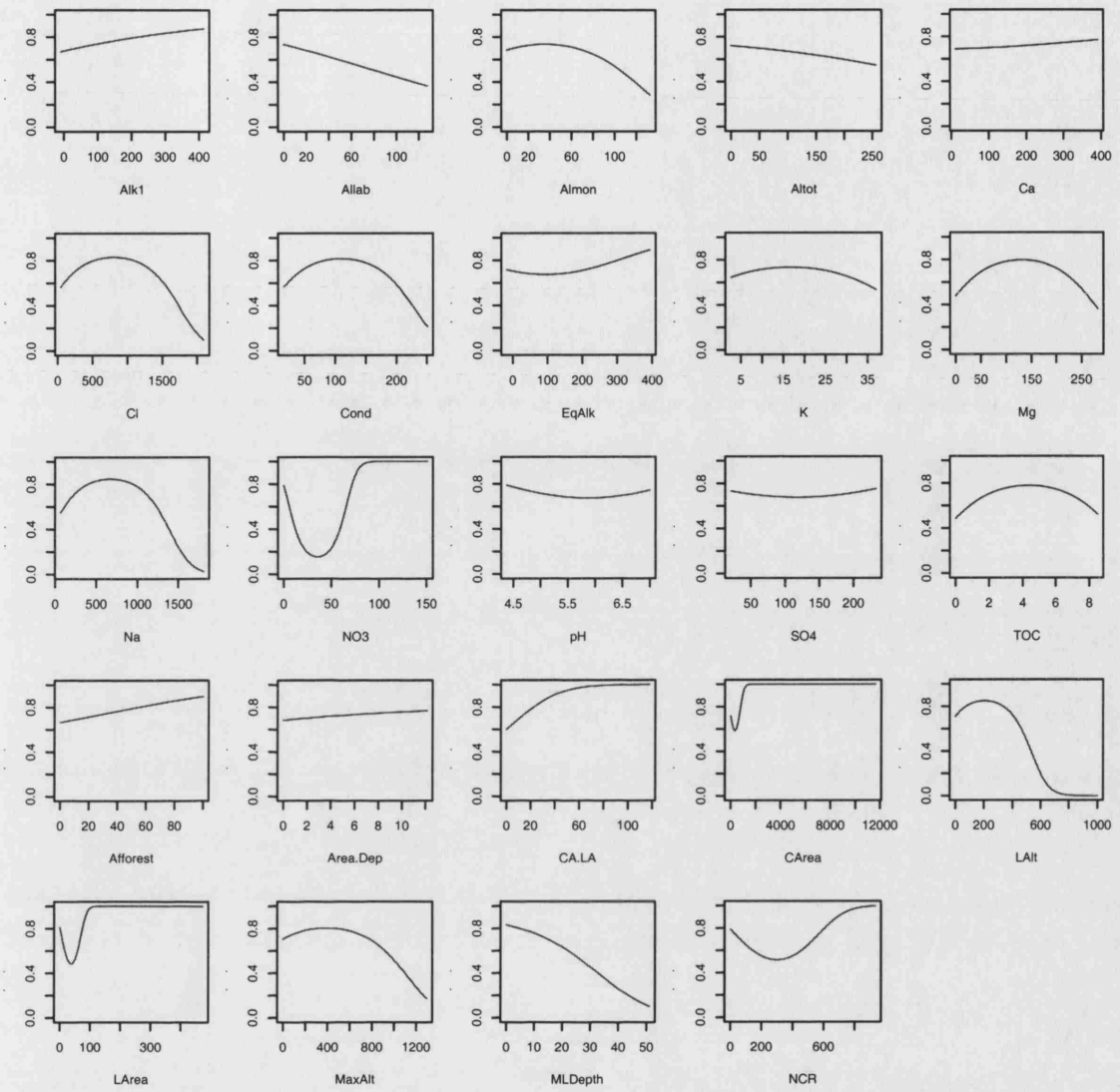


Figure 121: Generalised linear models of the response of Graptoleberis testudinaria to physicochemical parameters. The y-axis is the modelled probability of occurrence.

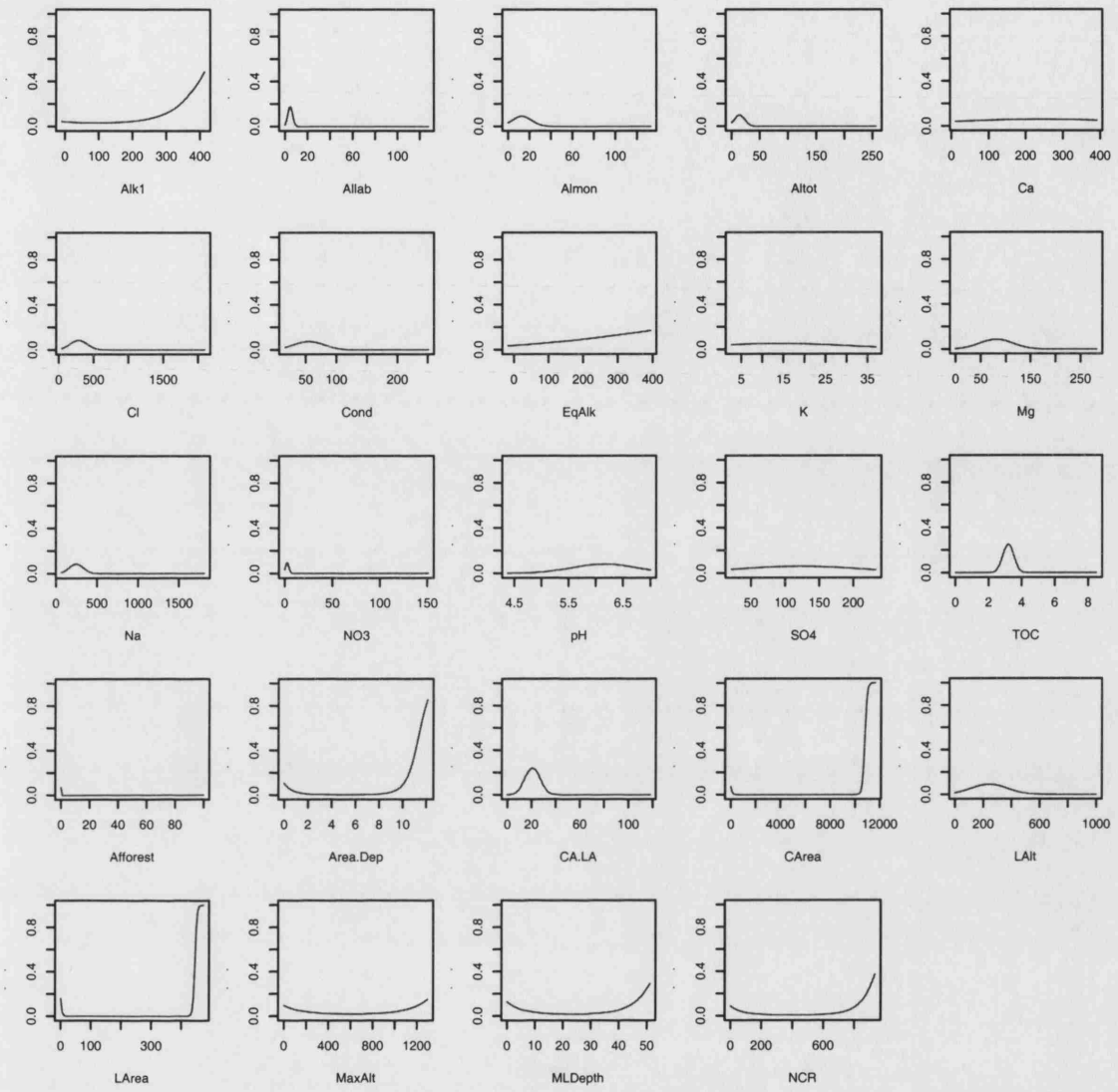


Figure 122: Generalised linear models of the response of Ilyocryptus sordidus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

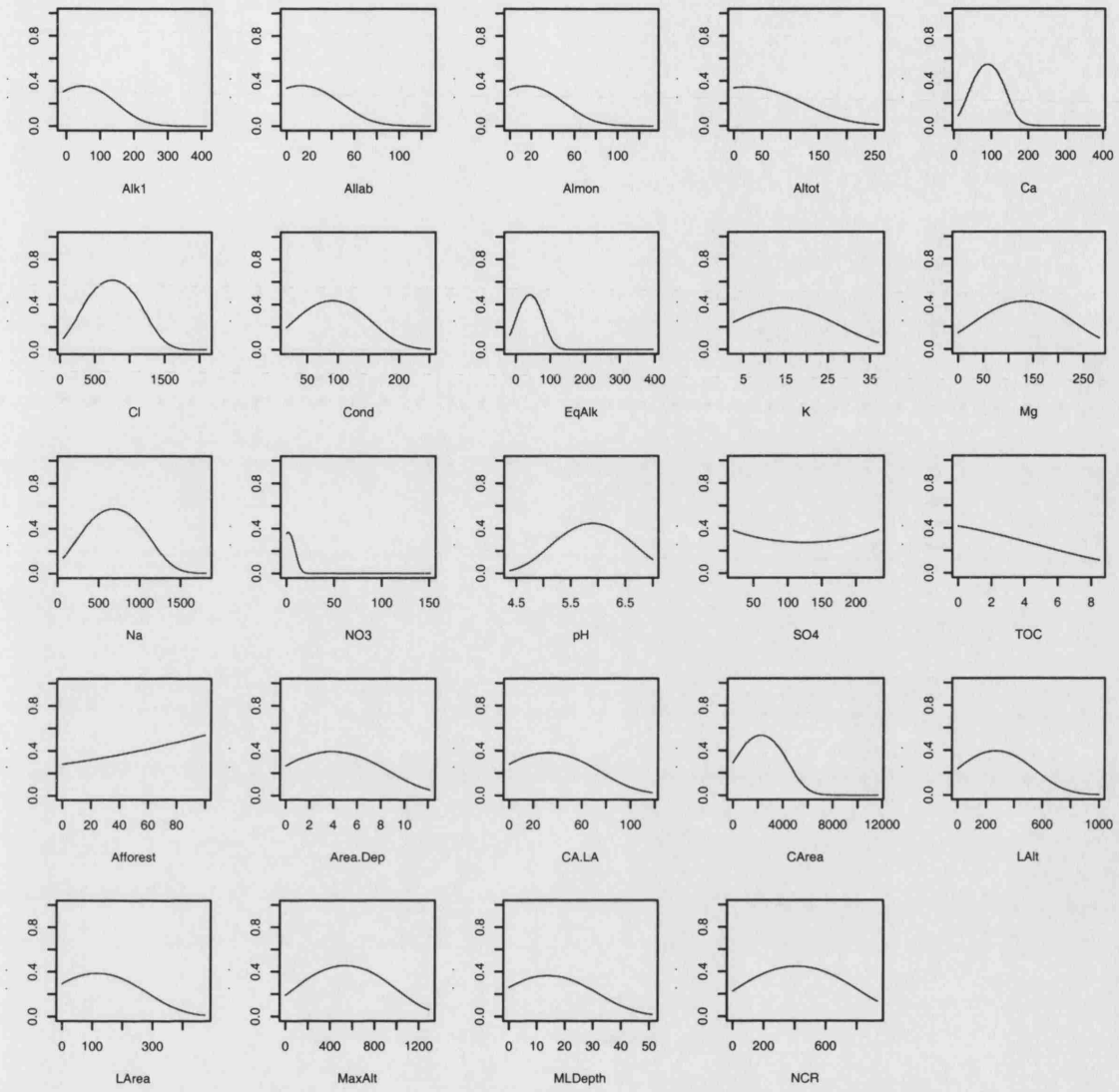


Figure 123: Generalised linear models of the response of large Alona spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

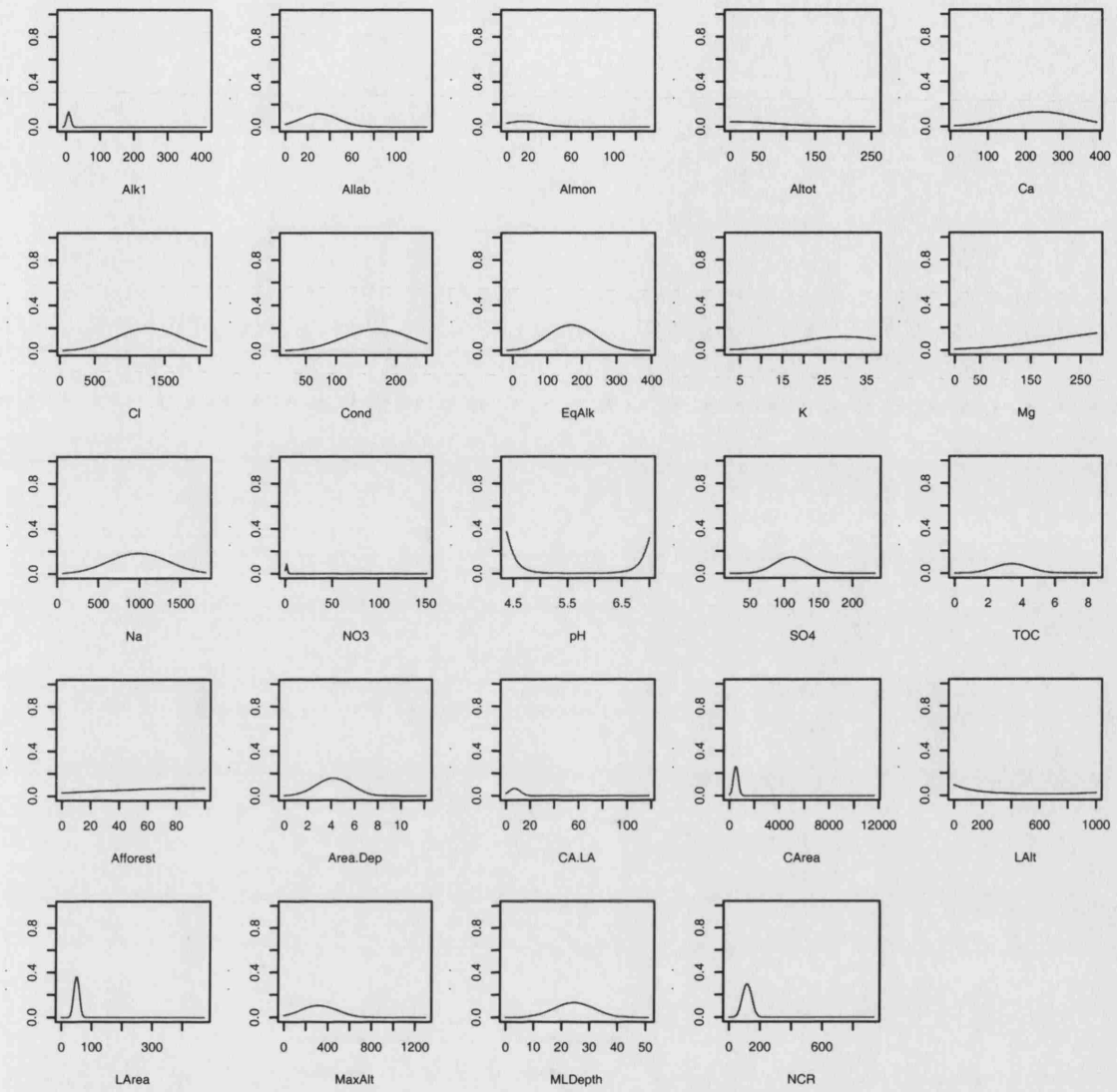


Figure 124: Generalised linear models of the response of Latona setifera to physicochemical parameters. The y-axis is the modelled probability of occurrence.

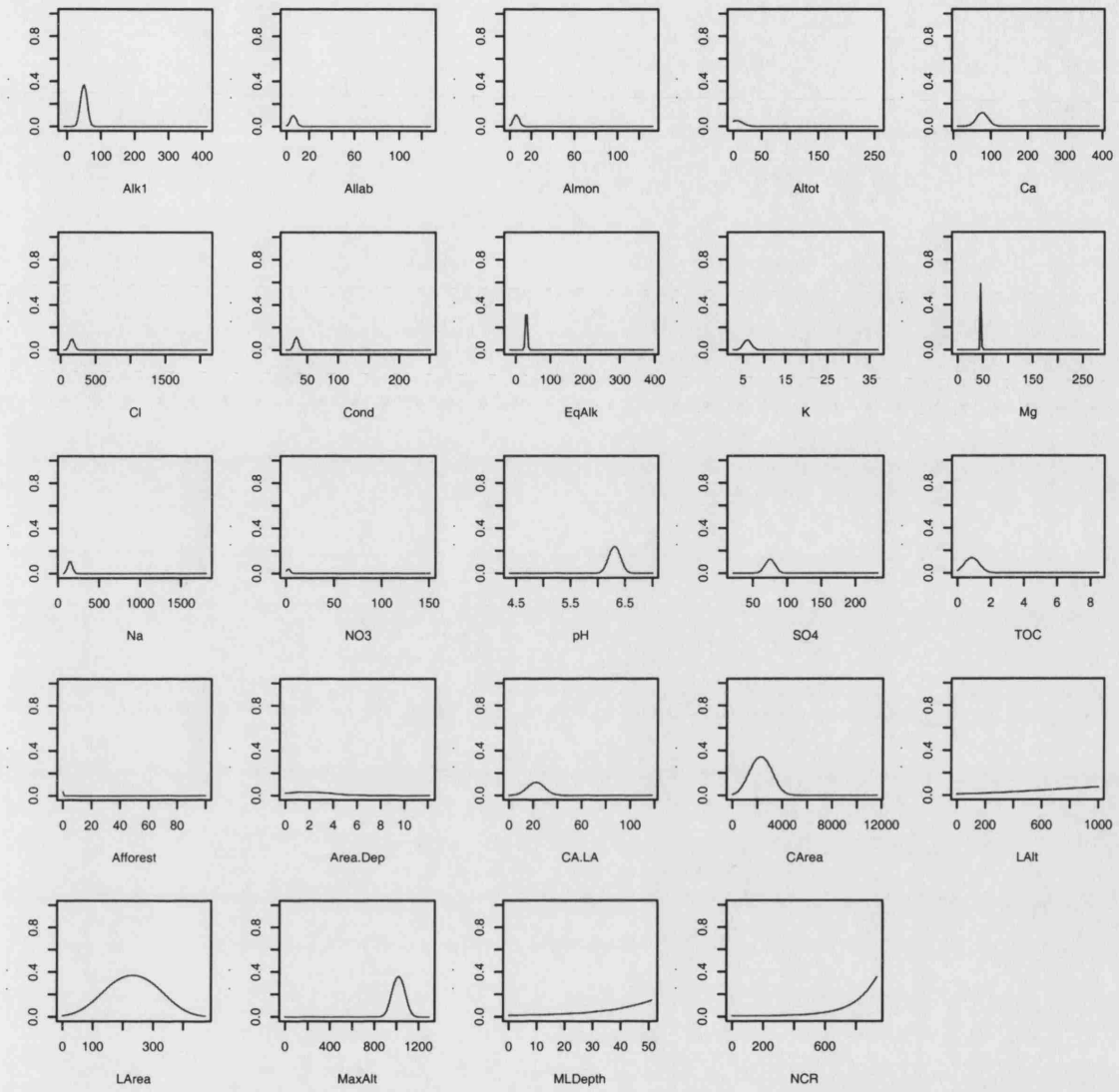


Figure 125: Generalised linear models of the response of Leydigia acanthocercoides to physicochemical parameters. The y-axis is the modelled probability of occurrence.

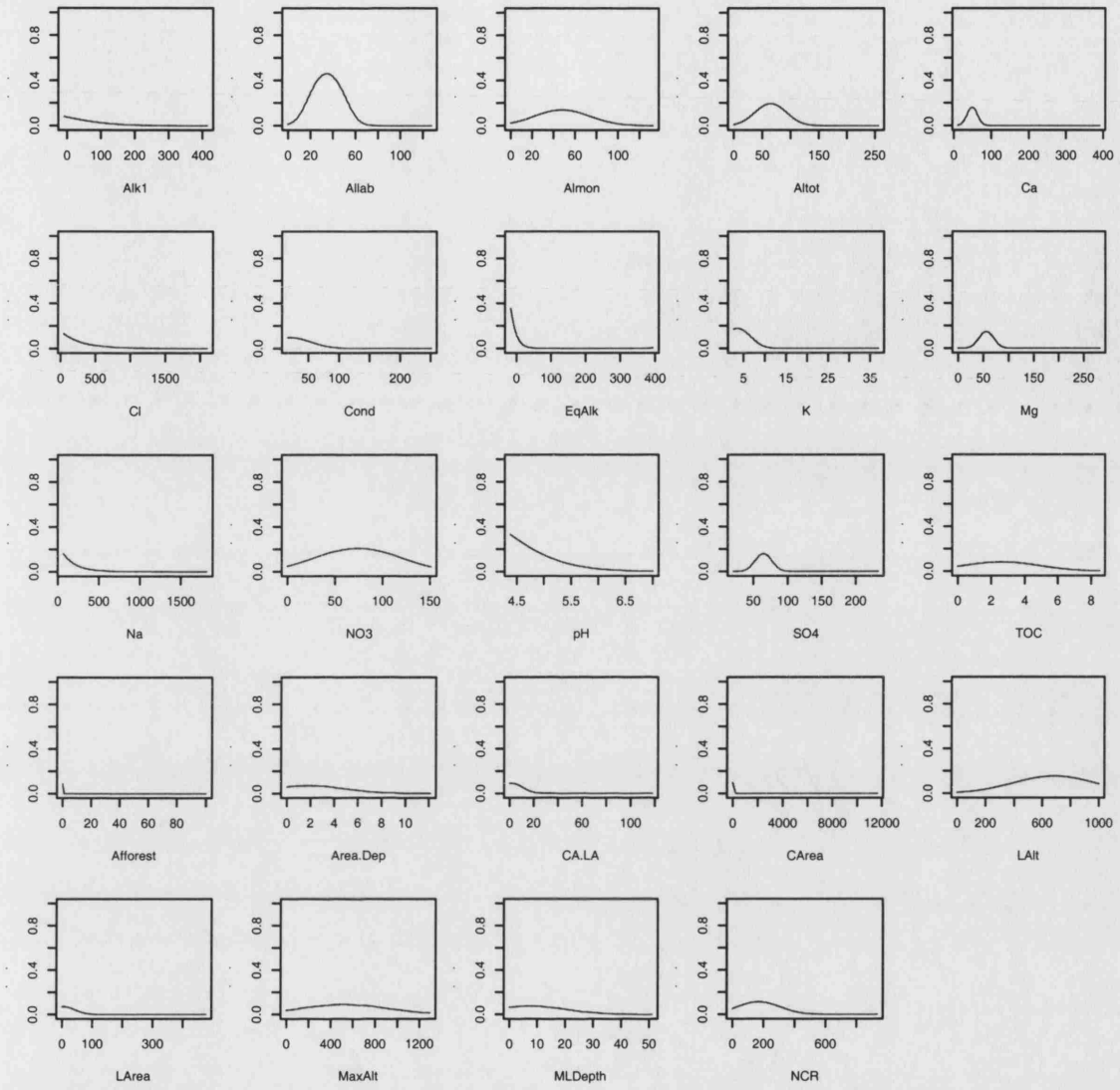


Figure 126: Generalised linear models of the response of Leydigia leydigi to physicochemical parameters. The y-axis is the modelled probability of occurrence.

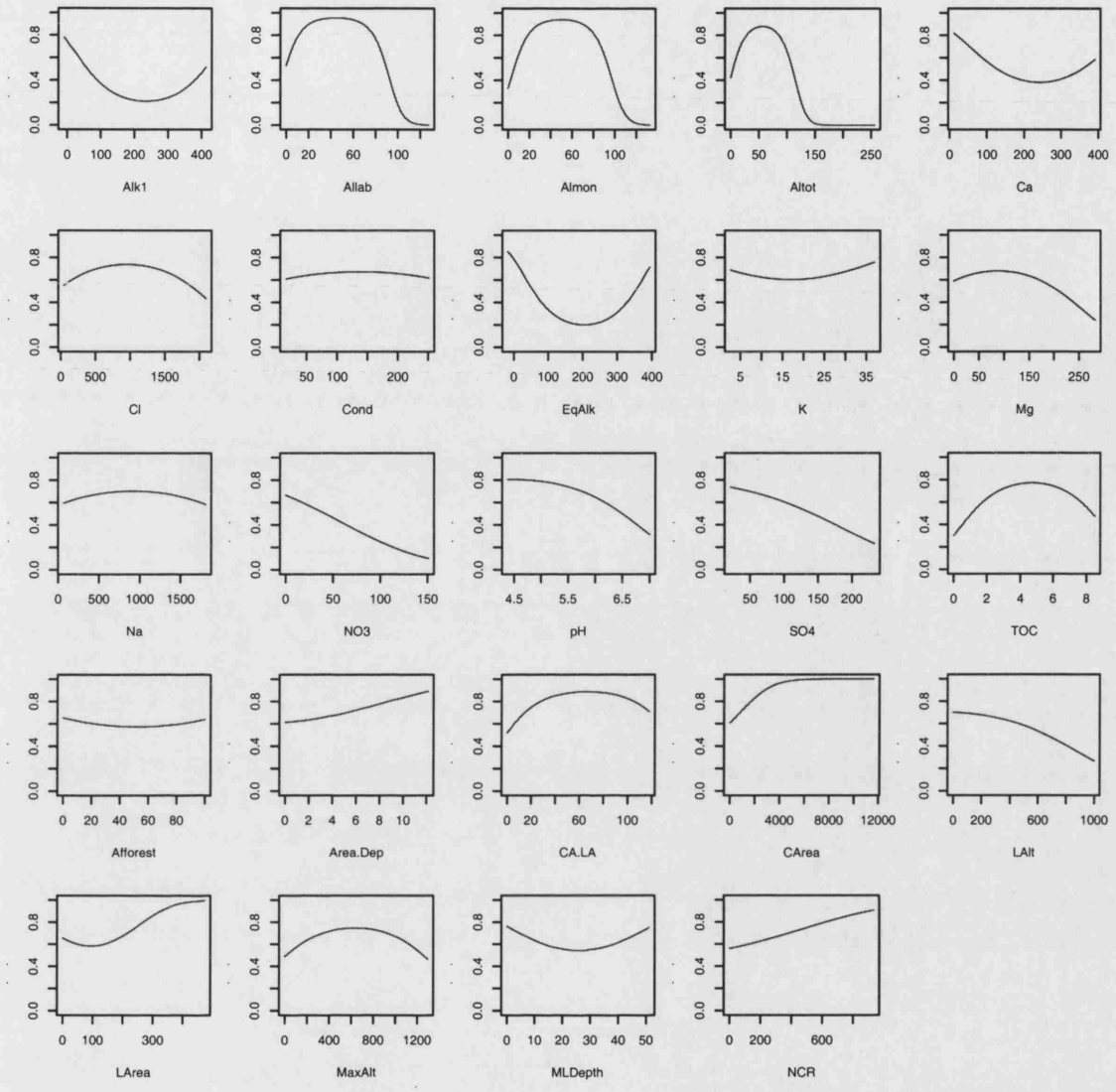


Figure 127: Generalised linear models of the response of Monospilus dispar to physicochemical parameters. The y-axis is the modelled probability of occurrence.

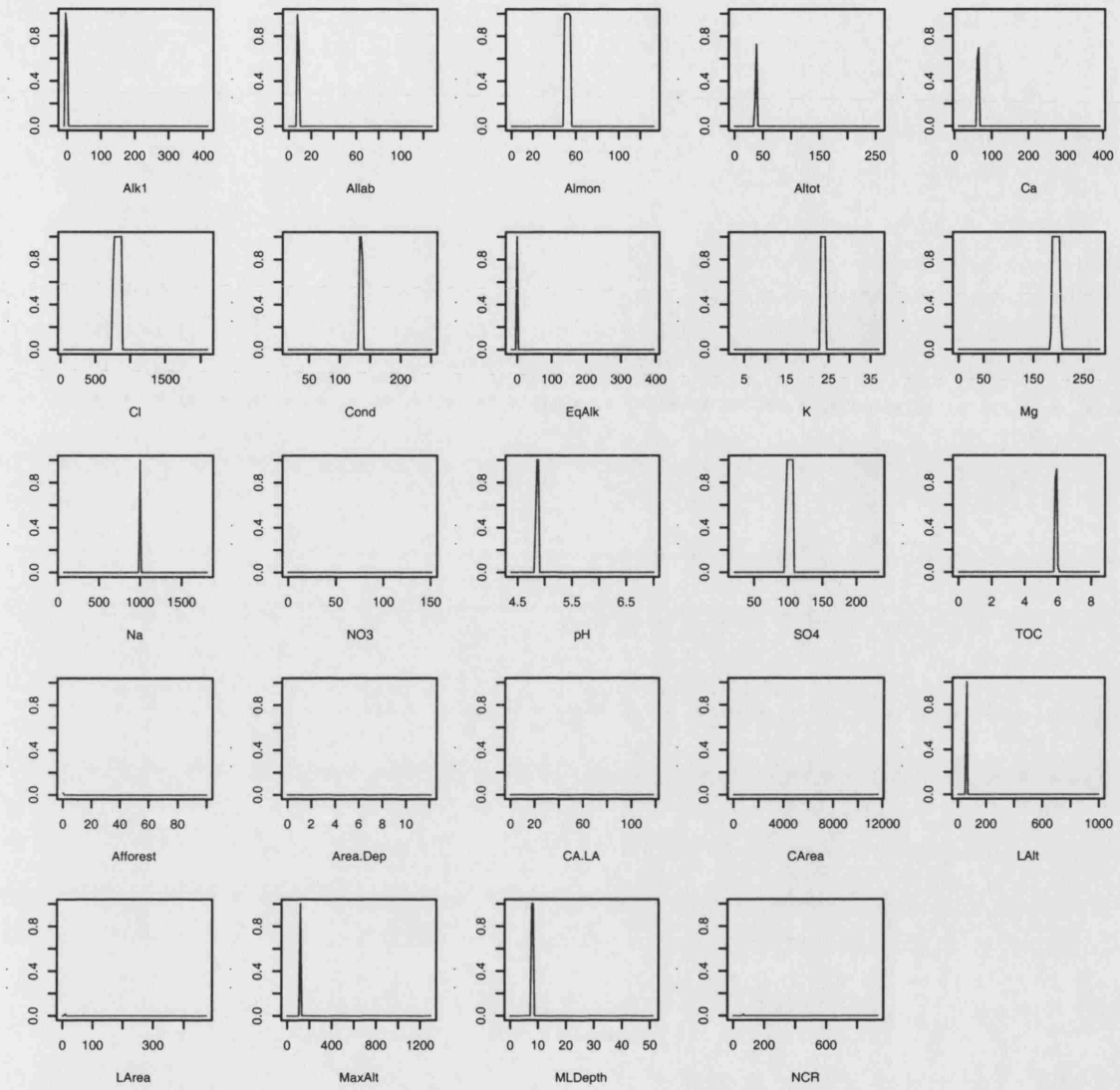


Figure 128: Generalised linear models of the response of Oxyurella tenuicaudis to physicochemical parameters. The y-axis is the modelled probability of occurrence.

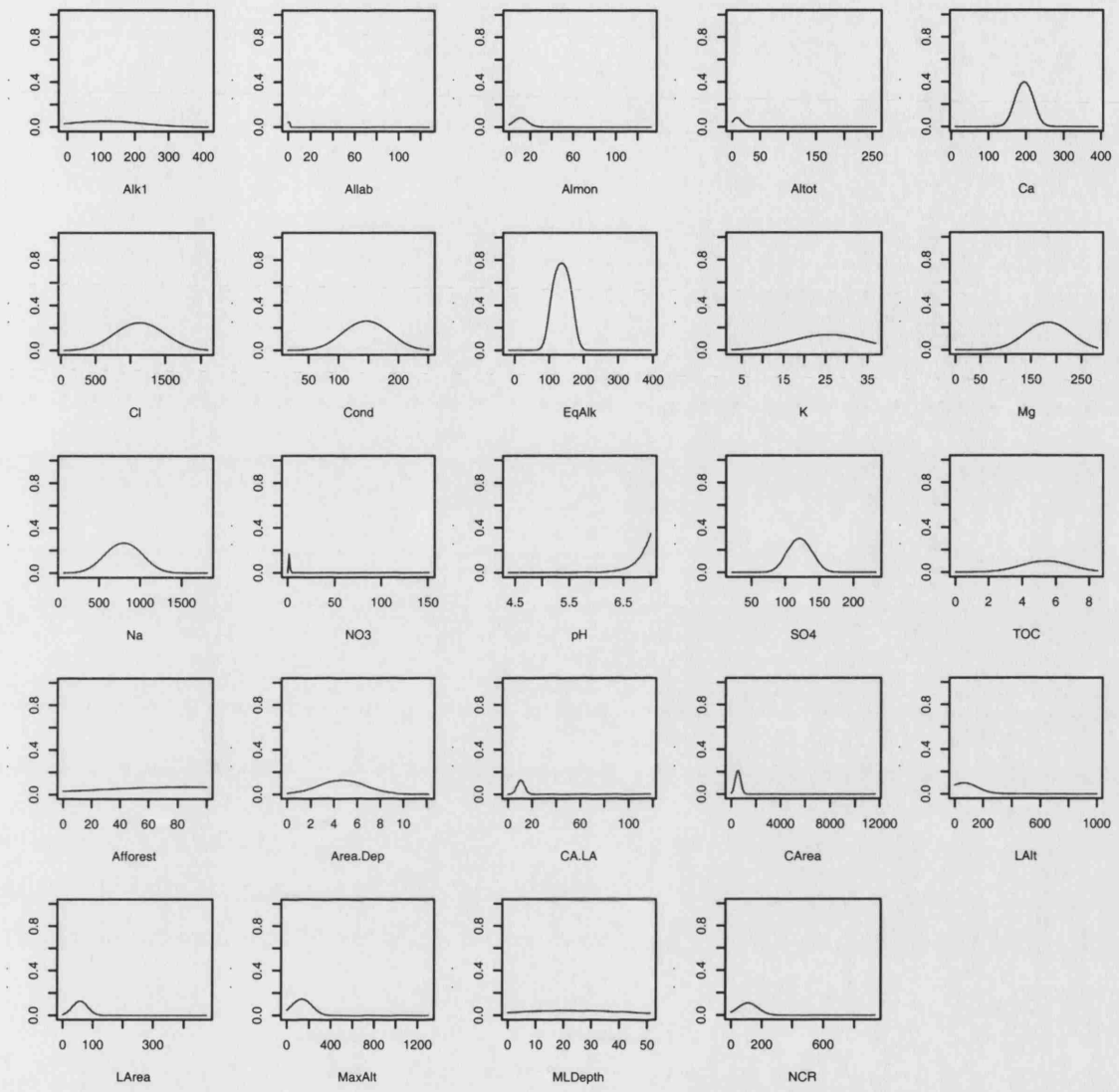


Figure 129: Generalised linear models of the response of Pleuroxus truncatus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

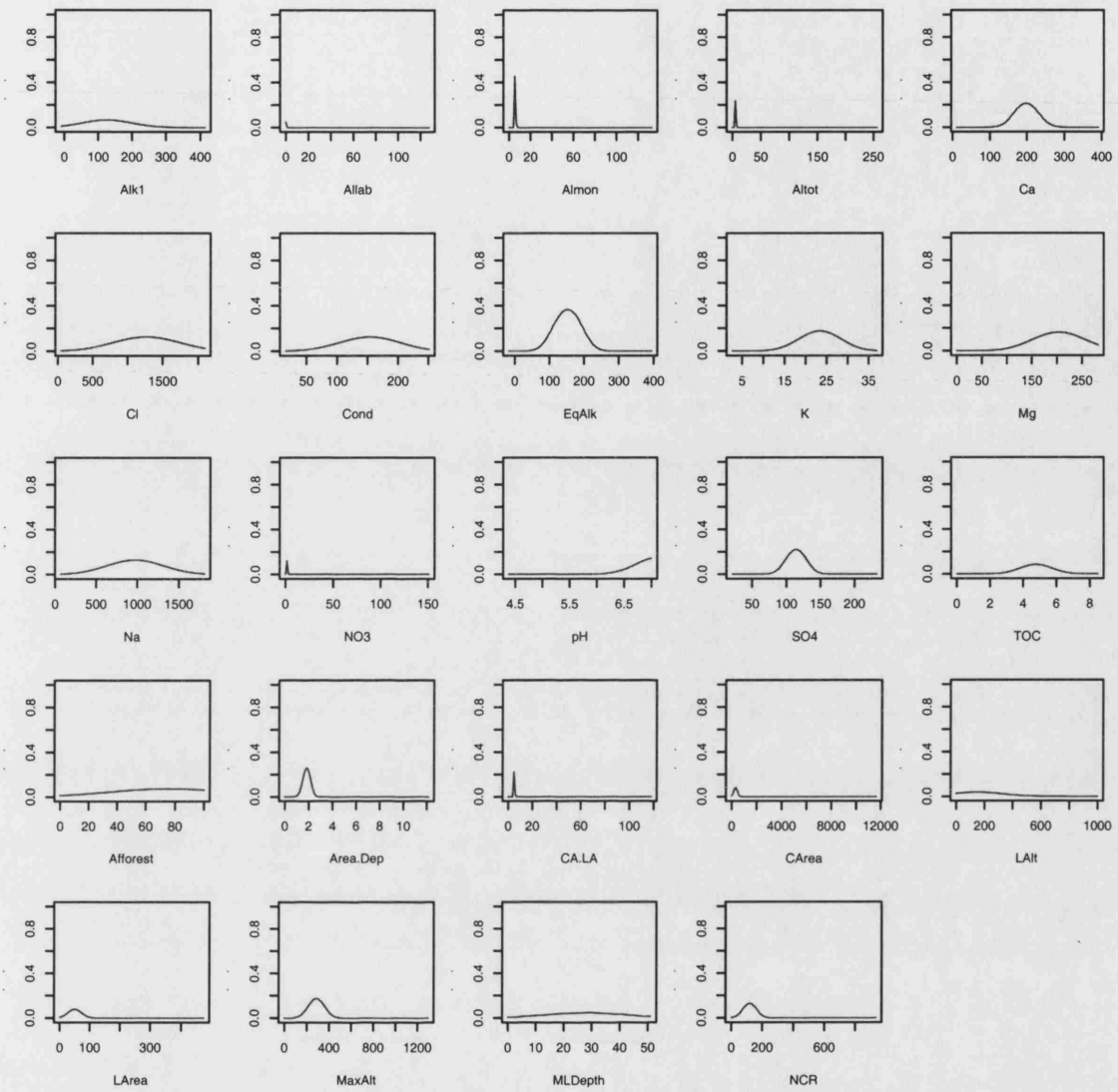


Figure 130: Generalised linear models of the response of Pleuroxus uncinatus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

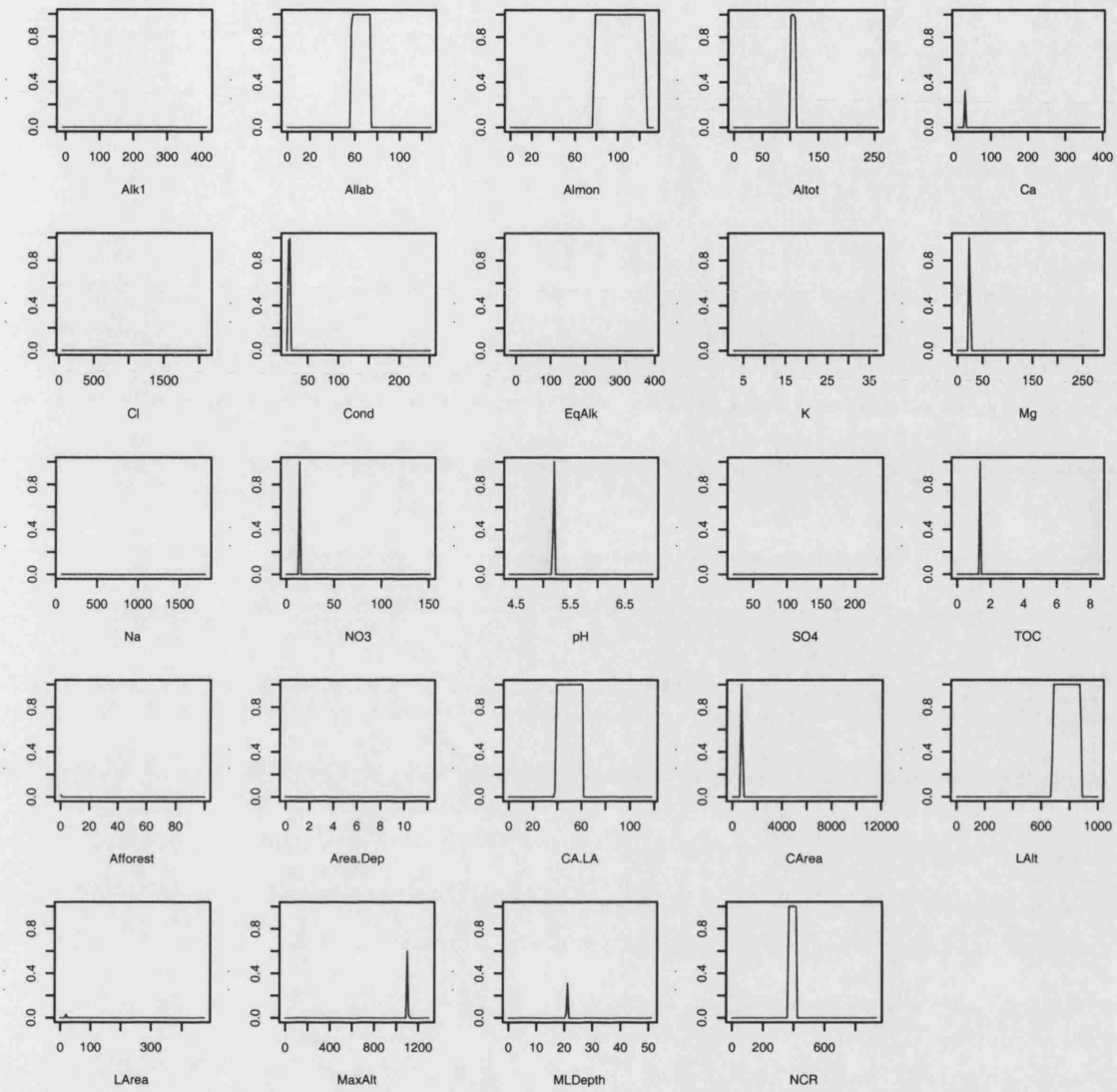


Figure 131: Generalised linear models of the response of Pseudochydorus globosus to physicochemical parameters. The y-axis is the modelled probability of occurrence.

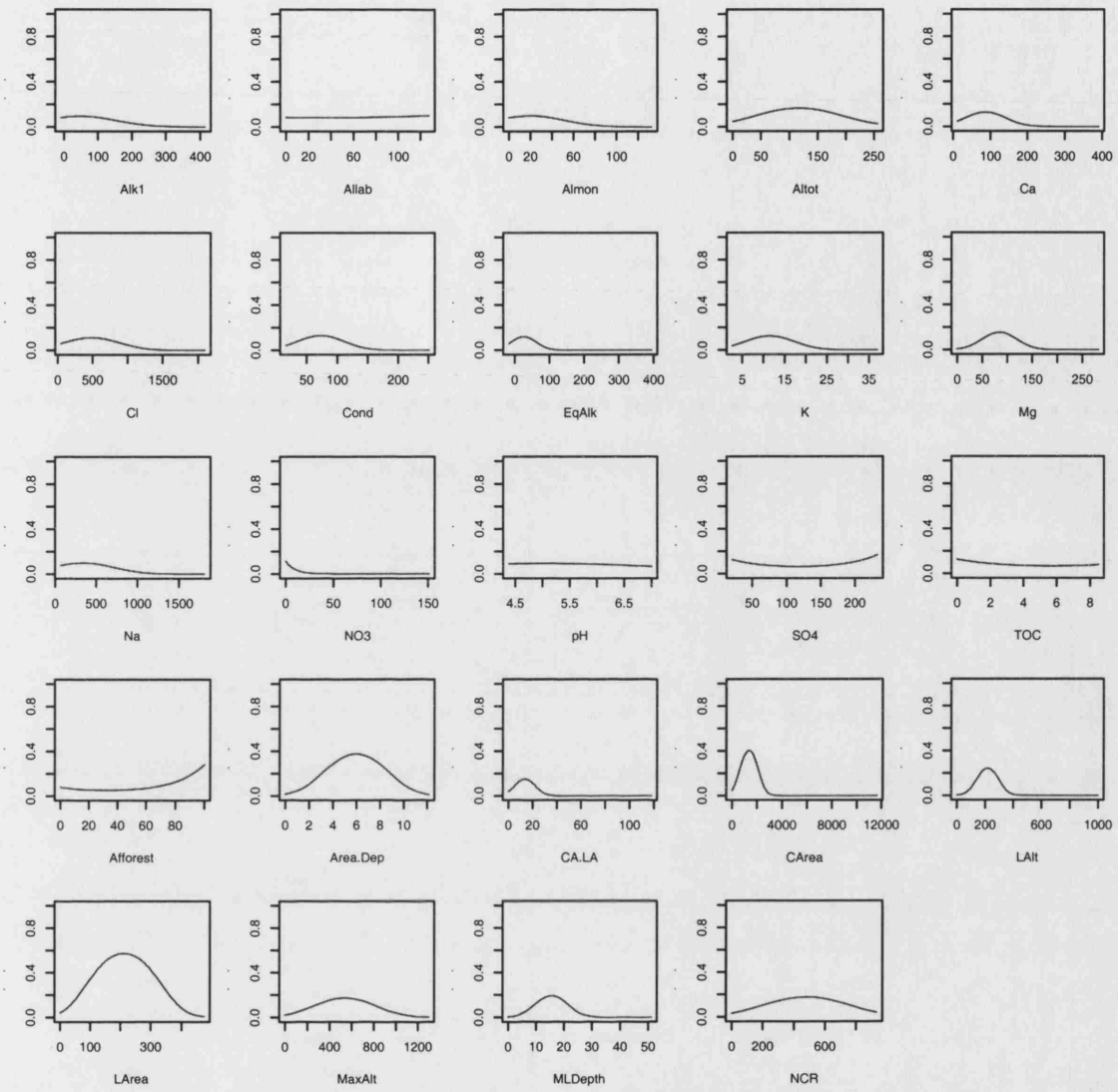


Figure 132: Generalised linear models of the response of Rhynchotalona falcata to physicochemical parameters. The y-axis is the modelled probability of occurrence.

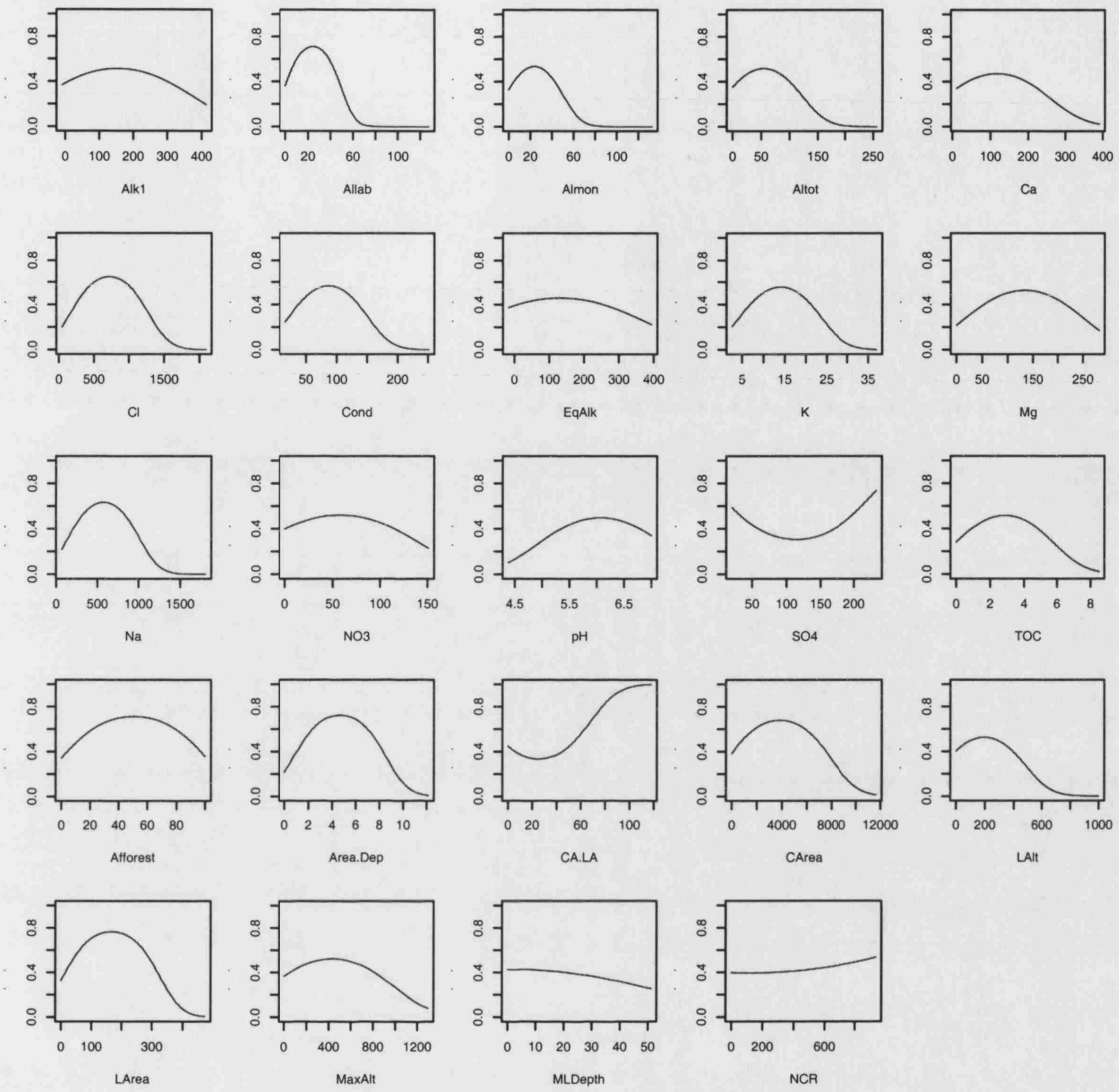


Figure 133: Generalised linear models of the response of Sida crystallina to physicochemical parameters. The y-axis is the modelled probability of occurrence.

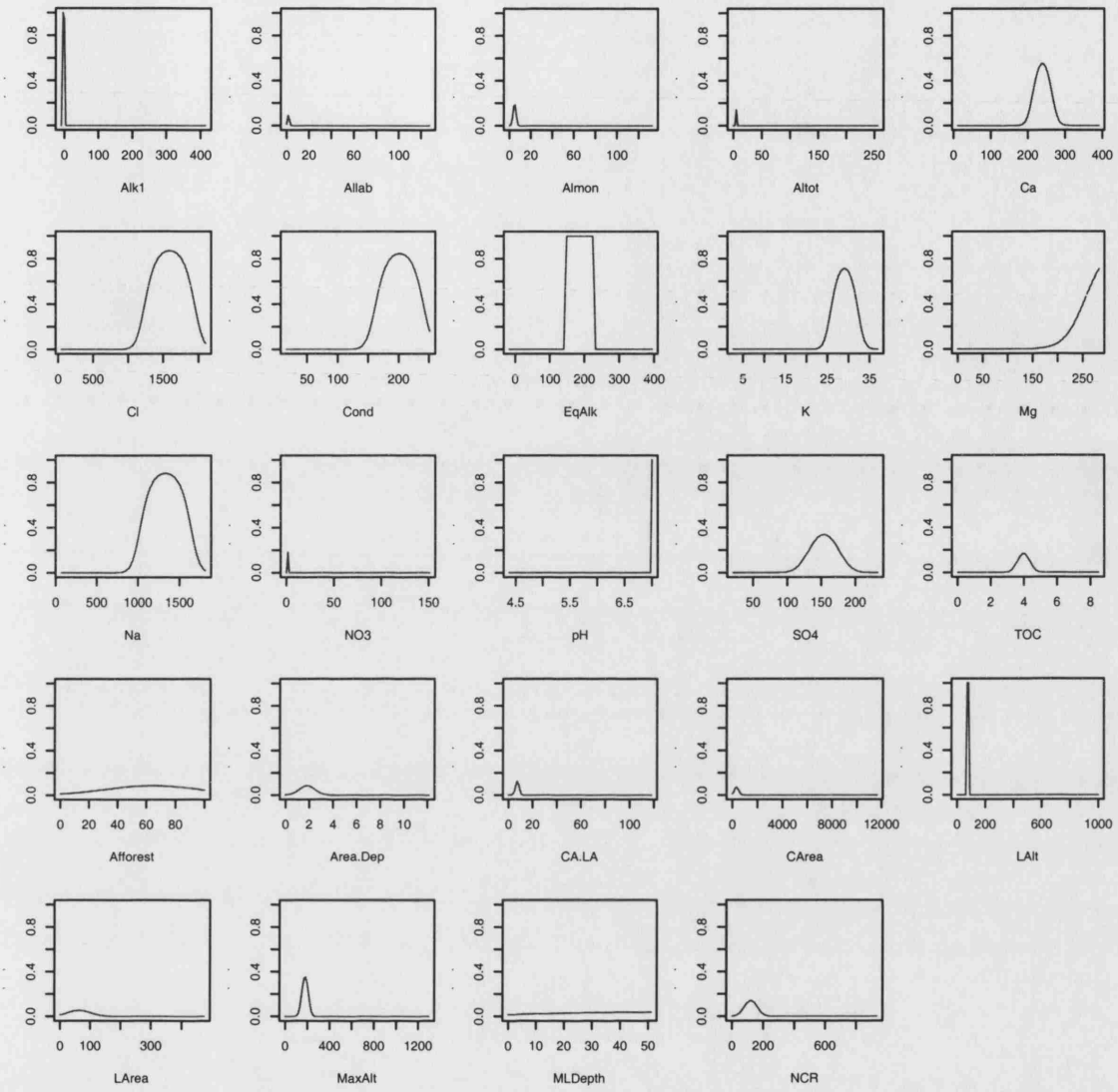


Figure 134: Generalised linear models of the response of Simocephalus spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

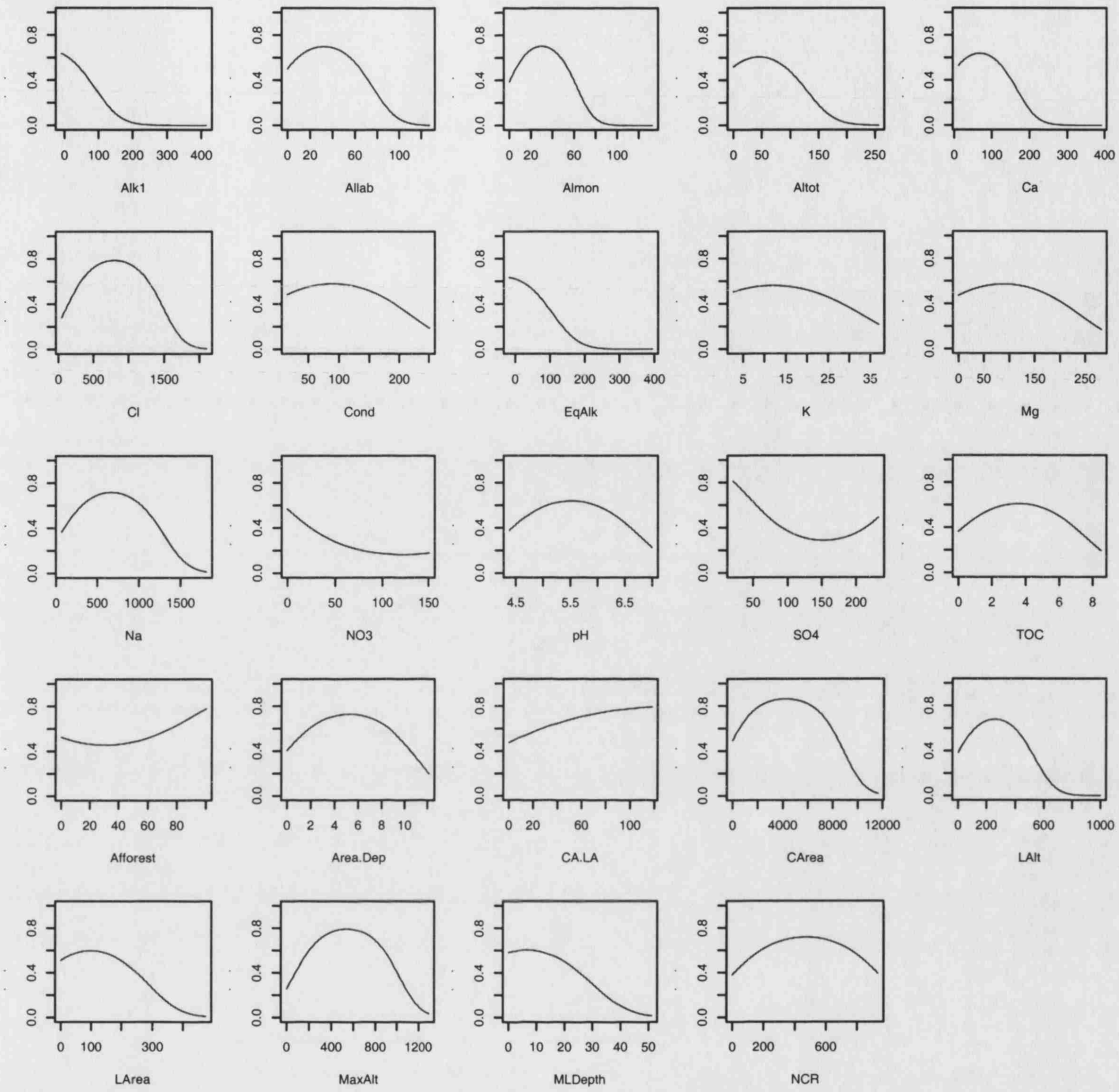


Figure 135: Generalised linear models of the response of small Alona spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.

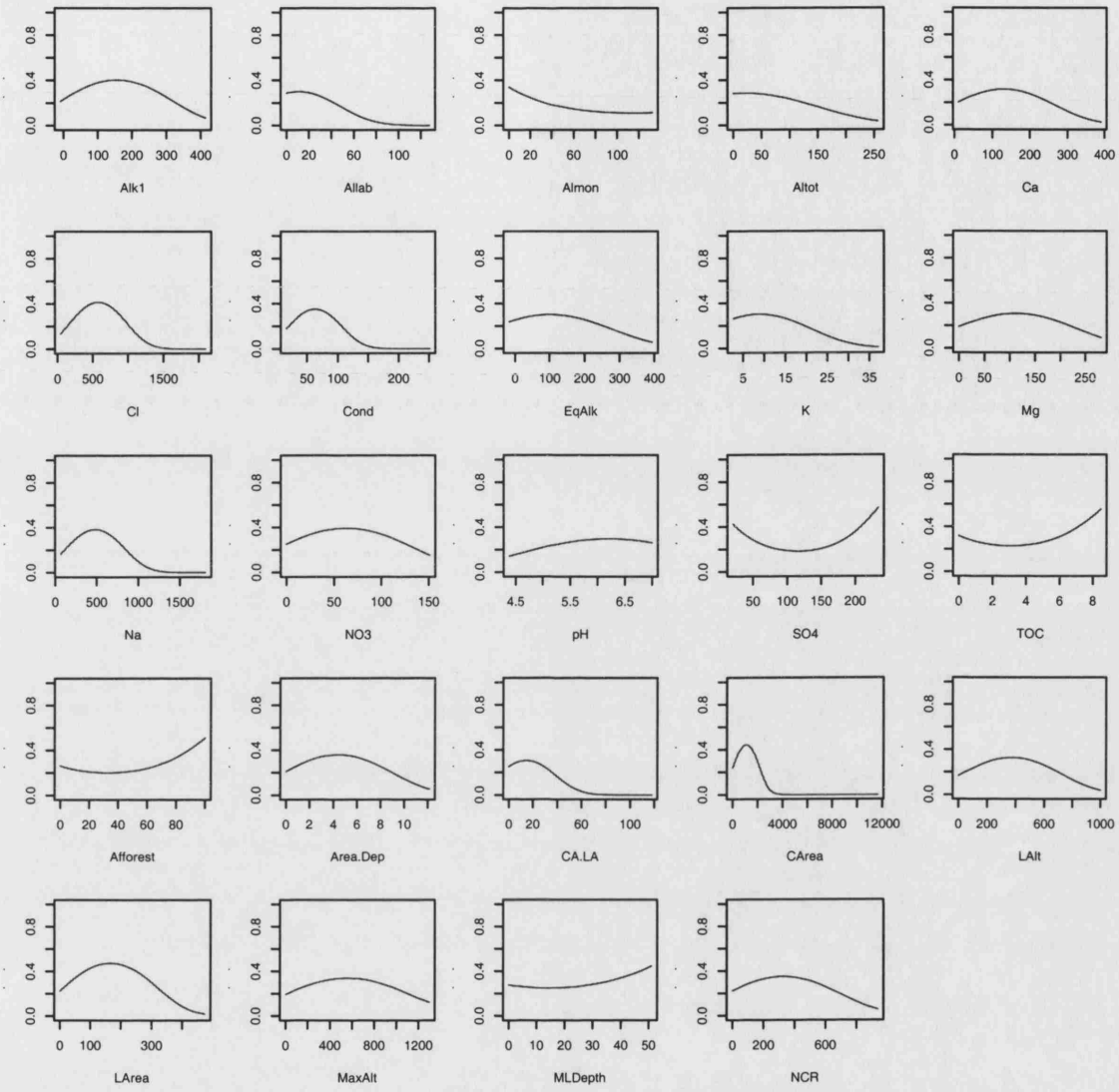


Figure 136: Generalised linear models of the response of Unknown spp. to physicochemical parameters. The y-axis is the modelled probability of occurrence.